



De la Génétique des Populations à la Génétique Quantitative

Biométrie

Claire BARIL
Jean-Claude BERGONZINI
U.R. AnIS

Septembre 1994

CIRAD-Forêt

Département Forestier du Centre de Coopération Internationale
en Recherche Agronomique pour le Développement

45 bis, avenue de la Belle-Gabrielle
94736 NOGENT-sur MARNE Cédex (FRANCE)

SOMMAIRE

	Pages
INTRODUCTION	1
1. CAS D'UN LOCUS	2
1.1 Modèles Génétiques et Statistiques (locus biallélique)	2
1.2 Principaux paramètres statistiques (locus biallélique)	4
1.3 Cas d'un locus multiallélique	8
1.4 Importance de la Dominance	8
2. MODELE A PLUSIEURS LOCI	10
2.1 En l'absence d'épistasie	10
2.2 En présence d'épistasie et d'équilibre de liaison	10
2.3 Le déséquilibre de liaison	12
3. NOTIONS D'APPARENTEMENT	15
3.1 Coefficient de simple parenté	15
3.2 Covariances entre apparentés (en régime panmictique)	16
4. EFFETS DES SYSTEMES DE REPRODUCTION	20
4.1 La consanguinité	20
4.2 Hybridation entre populations	22
4.3 Effet de la panmixie	23
5. GÉNOTYPE ET PHÉNOTYPE	25
5.1 Héritabilités	25
5.2 Covariances entre caractères	26
6. THÉOREME FONDAMENTAL DE LA SÉLECTION	28
6.1 Cas d'un locus multiallélique (Fisher)	28
6.2 Généralisation au cas multiloci (Ewens)	31
6.3 Passage au temps continu	32
7. LA GÉNÉTIQUE QUANTITATIVE	35
7.1 Modèle infinitésimal et approche Gaussienne	35
7.2 Lien avec le théorème de Fisher (Charnov 1989)	36
7.3 Généralisation aux cas d'appariements préférentiels	39
8. FORMALISATION DE TURELLI ET BARTON	42
ANNEXES	
QUELQUES OUVRAGES DE RÉFÉRENCE	

Au commencement était MENDEL ruminant ses pensées solitaires. Puis il dit : "Qu'il y ait des pois", et il leur dit : "Croissez et multipliez, différenciez-vous et assortissez-vous indépendamment". Ainsi firent-ils, et cela était bon. Puis advint que MENDEL rassembla ses pois et les sépara en graines rondes et ridées; il appela les rondes dominantes, et les ridées récessives, et cela était bon. Mais MENDEL vit alors qu'il y avait 450 pois ronds et 102 pois ridés. Cela n'était pas bon. Car la loi stipule qu'il doit y avoir trois ronds pour un ridé. Et MENDEL se dit en lui-même: "Gott in Himmel, c'est là l'oeuvre d'un ennemi qui aura semé des mauvais pois dans mon jardin à la faveur de la nuit". Et MENDEL pris d'un juste courroux, frappa sur la table et dit: "Eloignez-vous de moi, pois maudits et diaboliques, retournez dans ces ténèbres où vous serez dévorés par les rats et les souris!" Et il en fut ainsi; il ne resta plus que 300 pois ronds et 100 pois ridés, et cela était bon.

Excellent même. Et MENDEL le publia.

Anonyme 1972 'Peas on earth'

Hort. Sciences N7, p.5.

The familiar Mendelian ratios, which display the mechanism of inheritance, can be seen only when a gene difference at a single locus gives rise to a readily detectable difference in some such property of the organism. Quantitative differences, in so far as they are inherited, depend on genes whose effects are small in relation to the variation arising from other causes. Furthermore, quantitative differences are usually, though not necessarily always, influenced by gene differences at many loci. Consequently the individual genes, whether few or many, cannot be identified by their segregation; the Mendelian ratios are not displayed, and the methods of Mendelian analysis cannot be applied.

It is, nevertheless, a basic premiss of quantitative genetics that the inheritance of quantitative differences depends on genes subject to the same laws of transmission and having the same general properties as the genes whose transmission and properties are displayed by qualitative differences. Quantitative genetics is therefore an extension of Mendelian genetics, resting squarely on Mendelian principles as its foundation.

The methods of study in quantitative genetics differ from those employed in Mendelian genetics in two respects. In the first place, since ratios cannot be observed, single progenies are uninformative, and the unit of study must be extended to 'population', that is, larger groups of individuals comprising many progenies. And, in the second place, the nature of the quantitative differences to be studied requires the measurement, and not just the classification, of the individuals. The extension of Mendelian genetics into quantitative genetics may thus be made in two stages, the first introducing new concepts connected with the genetic properties of 'populations' and the second introducing concepts connected with the inheritance of measurements...

These two parts of the subject are often distinguished by different names, the first being referred to as 'population genetics' and the second as 'quantitative genetics' or 'biometrical genetics'.

'Introduction to QUANTITATIVE GENETICS'-
D. S. FALCONER, 1989

DE LA GÉNÉTIQUE DES POPULATIONS A LA GÉNÉTIQUE QUANTITATIVE

Claire Baril avec la collaboration de Jean-Claude Bergonzini

Introduction

La théorie de la sélection intègre deux types d'approches analytiques:

la **Génétique des populations** qui permet d'étudier des caractères déterminés par un ou deux gènes (variations discrètes),

la **Génétique Quantitative**, qui utilise les approximations issues des propriétés statistiques de la loi Normale en se basant sur les trois hypothèses suivantes:

- * les distributions des *phénotypes* et des *génotypes* sont Normales,
- * les caractères quantitatifs continus sont gouvernés par un nombre infini de gènes à effets infinitésimaux et tous indépendants 2 à 2 (le *Modèle infinitésimal* permet d'étendre au génotype entier les calculs exacts obtenus pour un seul gène),
- * les populations de départ sont infinies (pas de dérive).

Or, il est clair que ces hypothèses ne sont pas réalistes. En effet:

- . Si l'on peut considérer que la distribution des valeurs génotypiques suit une loi Normale, ceci n'est plus vrai sous l'effet de la sélection (troncature).
- . Les gènes sont portés par un nombre fini de chromosomes de taille finie, donc les gènes déterminant un caractère sont en nombre fini et seul un petit nombre d'entre eux peuvent être considérés comme indépendants.
- . En sélection artificielle la taille des populations est de plus en plus réduite, notamment avec l'avènement des nouvelles techniques de reproduction.

Les limites de ces deux approches étant posées, nous allons tenter de présenter les bases de la *théorie de la sélection*.

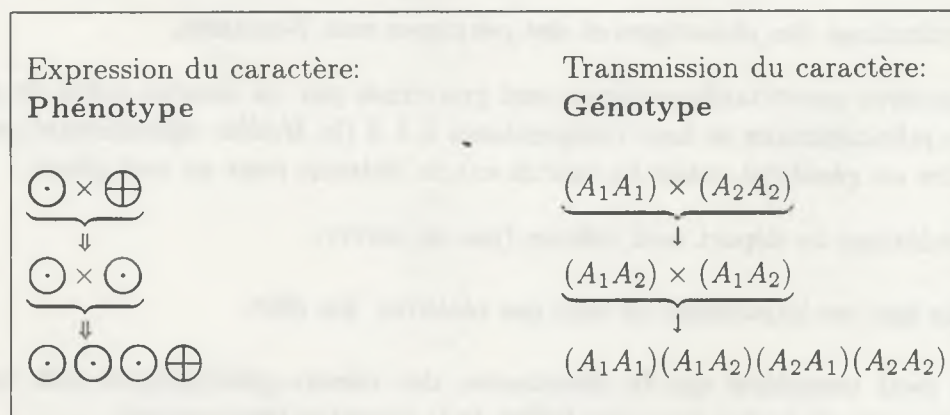
1 Cas d'un locus

1.1 Modèles Génétiques et Statistiques (locus biallélique)

Notre compréhension des mécanismes qui gouvernent l'*expression* et la *transmission* des caractères nous vient de la célèbre expérience que J. Mendel réalisa sur les pois (1865). Mendel disposait de pois lisses \odot et de pois ridés \oplus . Les ayant croisés, il obtint des pois lisses \odot qui recroisés entre eux donnèrent à nouveau des pois lisses et ridés, dans les proportions respectives de 3/4 et 1/4. Mendel fit alors trois hypothèses:

- 1 Au niveau individuel, le caractère est gouverné par un gène, support de l'expression de deux facteurs (le père et la mère), chacun de ces deux facteurs pouvant prendre deux modalités: les allèles A_1 et A_2 (plan factoriel 2^2).
- 2 Lorsque deux allèles différents sont présents chez un même individu, l'un peut être *dominant* A_1 (il détermine à lui seul l'expression du caractère), l'autre *récessif* A_2 .
- 3 Lors de la reproduction, un individu ne fournit qu'un seul des deux allèles dont il est pourvu.

Ces hypothèses permettent d'expliquer les résultats de Mendel à l'aide du schéma suivant:



Rappel:

Chez les individus diploïdes qui représentent la plupart des espèces cultivées, les *chromosomes* sont au nombre de $2n$. Ils constituent le support de l'information génétique. L'unité de transmission de cette information est le *locus*. A chacun des loci se situent 2 *gènes*, l'un provenant du père et l'autre de la mère, qui peuvent agir selon autant de modalités qu'il existe d'*allèles* (formes particulières que prend chaque gène, déterminant l'un des états possibles du caractère codé par ce gène).

Soit p et q les fréquences respectives des 2 allèles A_1 et A_2 ($p + q = 1$) dans la population étudiée, il existe quatre génotypes potentiels $A_1 A_1$, $A_1 A_2$, $A_2 A_1$ et $A_2 A_2$. On suppose dans la suite de l'exposé que le caractère étudié est quantifiable et que g_{11} , g_{12} , g_{21} et g_{22} sont les valeurs respectives associées à ces génotypes, on considère généralement qu'il n'y a pas d'effet lié au sexe ($A_1 A_2 = A_2 A_1$) et donc que $g_{12} = g_{21}$.

Sous l'hypothèse de **Hardy-Weinberg** qui postule que la fréquence d'un génotype est le produit des fréquences de ses allèles, les fréquences respectives des trois génotypes différents sont p^2 , $2pq$ et q^2 .

$$\text{H.W.} : (pA_1 + qA_2)^2 = p^2A_1A_1 + 2pqA_1A_2 + q^2A_2A_2$$

Pour comparer les valeurs génotypiques et tenir compte du comportement de l'hétérozygote par rapport aux homozygotes, on définit trois paramètres génétiques: la moyenne des homozygotes (c), l'effet de dominance (d) et l'effet additif de l'allèle A_1 (a).

$$\begin{cases} c = (g_{11} + g_{22})/2 \\ d = g_{12} - c \\ a = (g_{11} - g_{22})/2 = g_{11} - c = c - g_{22} \end{cases} \quad \begin{matrix} d > 0 \\ a > 0 \end{matrix}$$

Remarque:

On suppose le plus souvent que la dominance est positive. Cette attitude se justifie dans la mesure où les caractères sur lesquels porte la sélection sont mesurés par des variables qui prennent des valeurs d'autant plus grandes que le caractère est mieux adapté au milieu dans lequel s'exerce la sélection.

Différents **degrés de dominance** peuvent être envisagés:

si $d/a = 1$ \longrightarrow dominance complète

si $0 < d/a < 1$ \longrightarrow dominance partielle

si $d/a > 1$ \longrightarrow superdominance

$$\begin{array}{lll} g_{11} = 10 & g_{11} = 10 & g_{11} = 10 \\ g_{12} = g_{21} = 12 & g_{12} = g_{21} = 9 & g_{12} = g_{21} = 10 \\ g_{22} = 8 & g_{22} = 8 & g_{22} = 8 \\ \hookrightarrow c = 9 & c = 9 & c = 9 \\ d = 3 & d = 0 & d = 1 \\ a = 1 & a = 1 & a = 1 \\ d/a = 3 & d/a = 0 & d/a = 1 \end{array}$$

Ces degrés de dominance, fonctions de la valeur de l'hétérozygote (g_{12}) par rapport aux homozygotes (g_{11} et g_{22}) sont représentés sur le schéma suivant (Demarly, 1977):

g_{22}		$\frac{g_{11} + g_{22}}{2}$		g_{11}		
						g_{12}
Super-dom.	\uparrow	Dominance partielle	\uparrow	Dominance partielle	\uparrow	Super-dom.
	Dominance complète		Additivité parfaite		Dominance complète	

Le modèle proposé par Mendel met en relation la valeur génétique d'un individu avec deux facteurs (père i et mère j) à deux modalités ($1 \leq i \leq 2$ et $1 \leq j \leq 2$). On suppose que les deux facteurs sont identiques.

Dans le cadre de ces suppositions, la valeur associée à un génotype A_iA_j , peut s'écrire:

$$g_{ij} = \mu + \alpha_i + \alpha_j + \beta_{ij}$$

μ = Valeur moyenne de la population

α_i = Effet additif de l'allèle i ,

= avantage moyen pour un génotype d'avoir l'allèle $\begin{cases} A_1 \rightarrow \alpha_1 \\ A_2 \rightarrow \alpha_2 \end{cases}$

β_{ij} = Effet de dominance = écart aux valeurs additives

= interaction entre les allèles i et j ,

= effet spécifique du couple d'allèles.

A_1A_1	A_1A_2	A_2A_2
$2\alpha_1 + \beta_{11}$	$\alpha_1 + \alpha_2 + \beta_{12}$	$2\alpha_2 + \beta_{22}$
p^2	$2pq$	q^2

Si l'on prend un individu au hasard dans la population, sa valeur génotypique est une variable aléatoire G qui peut prendre les différentes valeurs suivantes:

$$g_{11} = \mu + 2\alpha_1 + \beta_{11}$$

$$g_{12} = \mu + \alpha_1 + \alpha_2 + \beta_{12}$$

$$g_{22} = \mu + 2\alpha_2 + \beta_{22}$$

De manière plus générale, la valeur génétique s'écrit: $G = \mu + A + D$ où A synthétise la partie additive (α_1, α_2) et D , les interactions entre allèles homologues ($\beta_{11}, \beta_{12}, \beta_{22}$).

1.2 Principaux paramètres statistiques (locus biallélique)

Considérons la population de distribution génique $pA_1 + qA_2$
et de distribution génotypique $p^2A_1A_1 + 2pqA_1A_2 + q^2A_2A_2$

sa valeur moyenne s'écrit: $\mu = p^2g_{11} + 2pqg_{12} + q^2g_{22} = c + 2pqd + (p - q)a$

et sa variance: $\sigma_G^2 = 2pq[a - d(p - q)]^2 + 4p^2q^2d^2$.

Ces valeurs sont aussi l'espérance et la variance de la variable G introduite ci-dessus.

Soit α_i la moyenne pondérée des génotypes ayant l'allèle A_i , les valeurs additives s'écrivent:

$$\alpha_1 = pg_{11} + qg_{12} - \mu = q[a - d(p - q)]$$

$$\alpha_2 = pg_{12} + qg_{22} - \mu = -p[a - d(p - q)] \quad \text{avec } \alpha_1p + \alpha_2q = 0$$

Les valeurs de dominance s'écrivent:

$$\beta_{11} = g_{11} - \mu - 2\alpha_1 = -2q^2d$$

$$\beta_{12} = g_{12} - \mu - \alpha_1 - \alpha_2 = 2pqd$$

$$\beta_{22} = g_{22} - \mu - 2\alpha_2 = -2p^2d \quad \text{avec } p\beta_{11} + q\beta_{12} = p\beta_{12} + q\beta_{22} = 0$$

On peut alors décomposer la variance génétique totale σ_G^2 de la manière suivante:

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2$$

$$\sigma_A^2 = p\alpha_1^2 + q\alpha_2^2 = 2pq[a - d(p - q)]^2$$

$$\sigma_D^2 = p^2\beta_{11}^2 + 2pq\beta_{12}^2 + q^2\beta_{22}^2 = 4p^2q^2d^2$$

ce qui signifie que les variables A et D introduites précédemment sont indépendantes.

Remarque:

On a exprimé les différentes variances en fonction des paramètres d et a qui ne dépendent pas des pondérations retenues pour calculer la décomposition des g_{ij} , ce qui facilite l'interprétation.

Les deux séries de figures qui suivent illustrent l'importance relative des variances d'additivité (notée $V(A)$) et de dominance (notée $V(D)$) en fonction d'une part de la fréquence de l'allèle dominant A_1 (notée p), et d'autre part du degré de dominance (noté d/a).

La valeur de a a été arbitrairement fixée à 1.

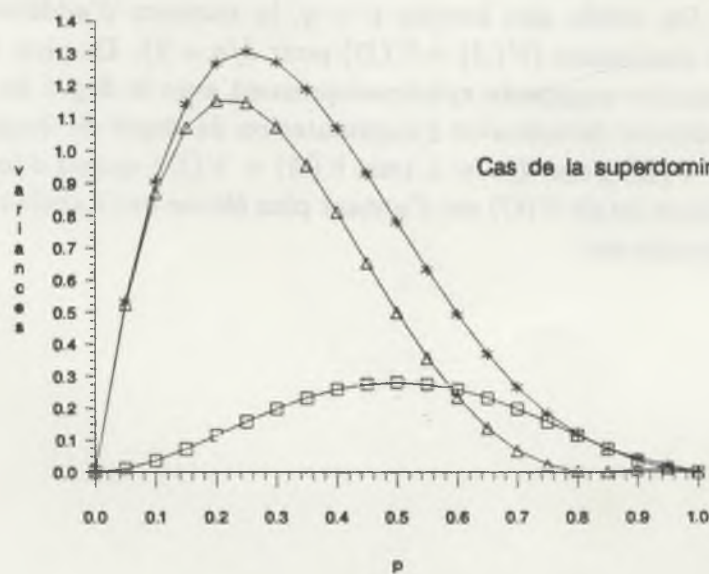
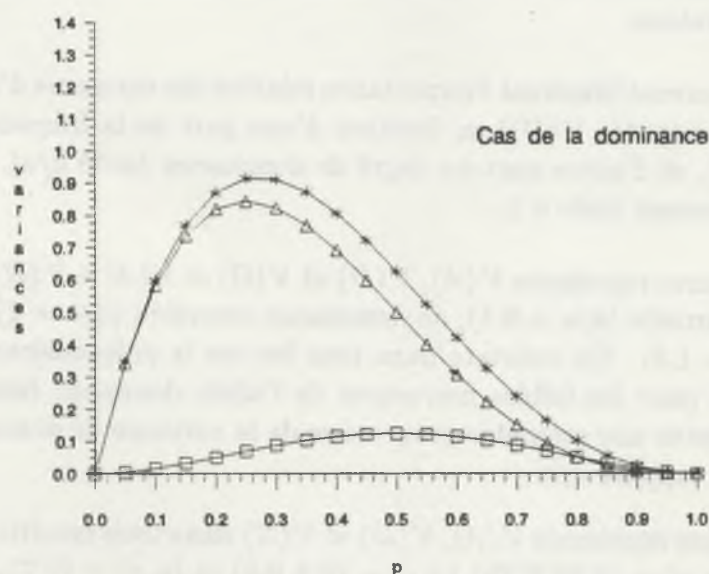
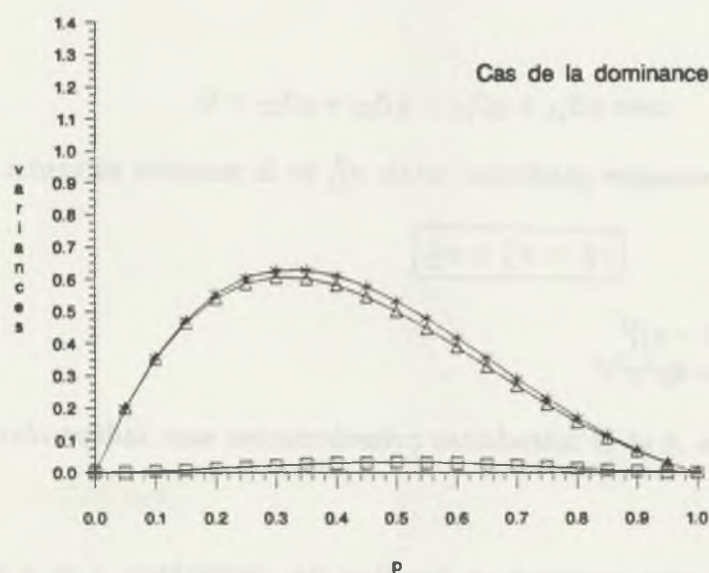
- * La première série de figures représente $V(A)$, $V(D)$ et $V(G) = V(A) + V(D)$ dans les cas de dominance partielle ($d/a = 0.5$), de dominance complète ($d/a = 1$) et de superdominance ($d/a = 1.5$). On constate dans tous les cas la prépondérance de la variance d'additivité pour les faibles fréquences de l'allèle dominant favorable A_1 . Par ailleurs, on observe une remontée progressive de la variance de dominance avec l'augmentation du rapport d/a .
- * La seconde série de figures représente $V(A)$, $V(D)$ et $V(G)$ dans trois conditions de fréquences alléliques: $(p, q) = (0.25, 0.75)$, $(p, q) = (0.5, 0.5)$ et $(p, q) = (0.75, 0.25)$. Les résultats obtenus viennent confirmer les observations faites à partir de la première série de figures. On vérifie que lorsque $p = q$, la variance d'additivité ne dépend pas du degré de dominance ($V(A) = V(D)$ pour $d/a = 2$). De plus, tandis que la variance de dominance augmente systématiquement avec le degré de dominance, la variance d'additivité décroît avec l'augmentation du degré de dominance lorsque $p > q$ ($V(A) = V(D)$ pour $d/a = 1$ puis $V(A) < V(D)$ quand $d/a > 1$). Enfin, la variance génétique totale $V(G)$ est d'autant plus élevée que l'allèle récessif défavorable A_2 est prépondérant.

$$c = (g_{11} + g_{22})/2$$

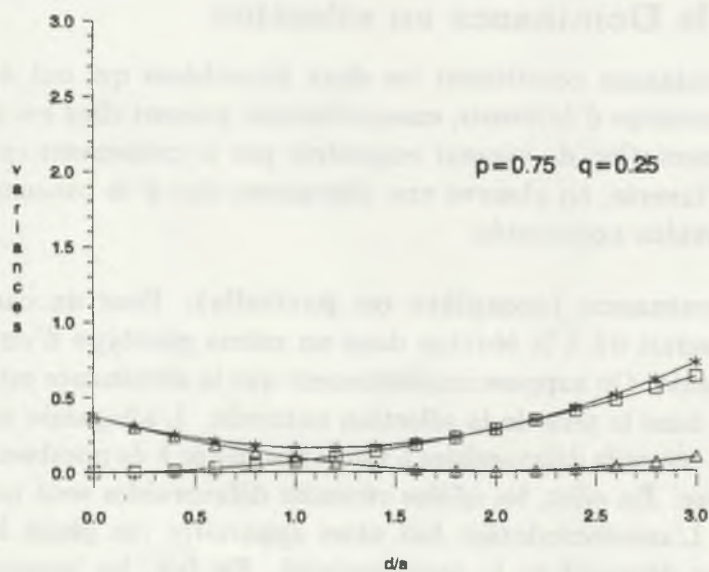
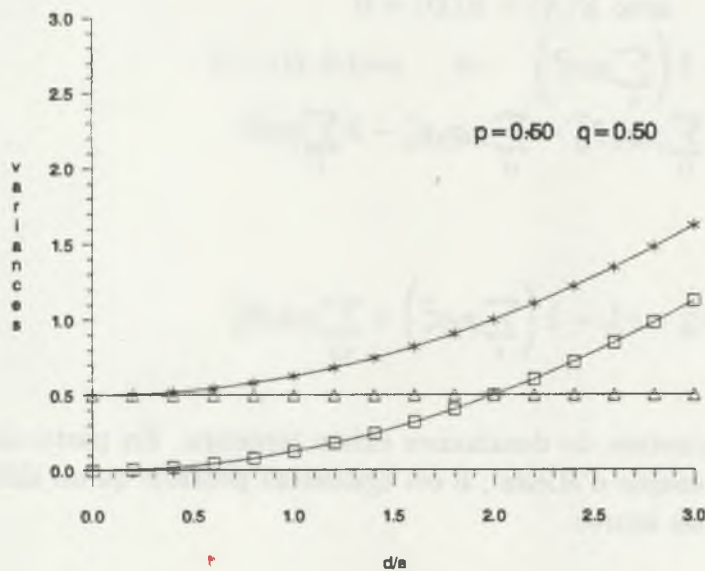
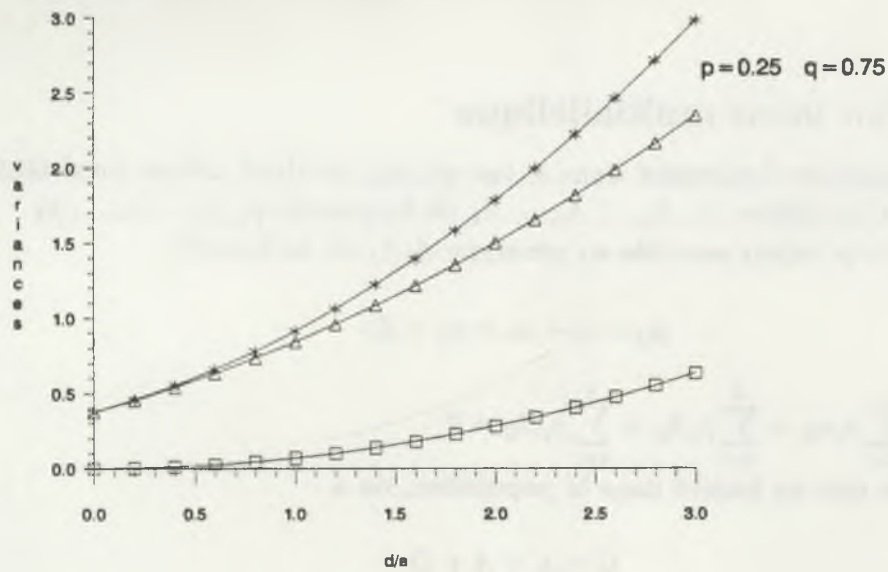
$$d = g_{12} - c$$

$$a = g_{11} - c$$

Importance relative des variances d'additivité
et de dominance dans le cas biallélique en fonction de p



$V(A)$ = triangle $V(D)$ = carre $V(G)$ = étoile



$V(A)$ = triangle $V(D)$ = carre $V(G)$ = étoile

1.3 Cas d'un locus multiallèlique

Le modèle se généralise facilement dans le cas où plus de deux allèles coexistent à un même locus. Soit les allèles $A_1, A_2, \dots, A_i, \dots, A_k$ de fréquences $p_1, p_2, \dots, p_i, \dots, p_k$.

La formulation de la valeur associée au génotype $A_i A_j$ est inchangée:

$$g_{ij} = \mu + \alpha_i + \alpha_j + \beta_{ij}$$

$$\text{avec } \sum_{i=1}^k p_i \alpha_i = \sum_{j=1}^k p_j \alpha_j = \sum_{i=1}^k p_i \beta_{ij} = \sum_{j=1}^k p_j \beta_{ij} = 0$$

Pour un individu pris au hasard dans la population, on a :

$$G = \mu + A + D$$

$$\text{avec } E(A) = E(D) = 0$$

$$\sigma_A^2 = 2 \left(\sum_i p_i \alpha_i^2 \right) \quad \text{et} \quad \text{cov}(A, D) = 0$$

$$\sigma_D^2 = \sum_{ij} p_i p_j \beta_{ij}^2 = \sum_{ij} p_i p_j g_{ij}^2 - 2 \sum_i p_i \alpha_i^2$$

La variance génétique s'écrit:

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 = 2 \left(\sum_i p_i \alpha_i^2 \right) + \sum_{i,j} p_i p_j \beta_{ij}^2$$

Dans le cas multiallèlique, la notion de dominance existe toujours. En particulier, on peut l'examiner pour chaque couple d'allèles ; il est également possible qu'un allèle soit dominant par rapport à tous les autres.

1.4 Importance de la Dominance en sélection

La dominance et la superdominance constituent les deux hypothèses qui ont été formulées pour expliquer le phénomène d'hétérosis, essentiellement présent chez les espèces allogames. Il s'agit de l'augmentation de vigueur engendrée par le croisement entre individus non apparentés. A l'inverse, on observe une *dépression due à la consanguinité* après reproduction entre individus apparentés.

Hypothèse de la dominance (complète ou partielle): Pour un caractère quantitatif, l'hétérosis serait dû à la réunion dans un même génotype d'un grand nombre d'allèles dominants. On suppose implicitement que la dominance est favorable ($d > 0$), ce qui va dans le sens de la sélection naturelle. L'allogamie autorise l'accumulation d'allèles récessifs défavorables à faible fréquence à de nombreux loci: c'est le *fardeau génétique*. En effet, les allèles récessifs défavorables sont masqués à l'état hétérozygote. L'autofécondation fait alors apparaître ces gènes à l'état homozygote, d'où l'effet dépressif de la consanguinité. En fait, les homozygotes

construits avec des allèles récessifs défavorables mettent en évidence le *fardeau génétique* et les homozygotes construits avec des allèles dominants favorables permettent de fixer l'hétérosis. Sous l'hypothèse de dominance complète, la **fixation de l'hétérosis** à long terme est envisageable, ce qui influe naturellement sur le choix d'une stratégie de sélection (lignées ou hybrides).

Hypothèse de la superdominance: La supériorité de l'hybride par rapport aux parents viendrait de la supériorité de l'état hétérozygote à certains loci. C'est la combinaison de deux gènes à l'état hétérozygote qui entraînerait une potentialité supérieure à celle des homozygotes. En fait, la superdominance apparaît souvent comme un phénomène marginal: on parle de *superdominance marginale*. L'hétérozygote présente un comportement plus stable sur l'ensemble des milieux ou sur l'ensemble de la vie d'une plante, sans jamais avoir été avantagé dans des conditions précises. La superdominance est alors la conséquence de l'accumulation de simple dominance dans différents milieux. Par ailleurs, la superdominance au niveau d'un caractère complexe peut être le résultat de gènes à effets *pléiotropiques* (la pléiotropie est l'action simultanée d'un gène sur plusieurs caractères). Sous l'hypothèse de superdominance: **l'hétérosis est infixable**, et par conséquent les variétés hybrides resteront toujours justifiées. Par ailleurs, la multiplication végétative ou la multiplication par *apomixie* (développement d'une graine sans fécondation) permettra de reproduire à l'identique les meilleures combinaisons hétérozygotes par voie non sexuée (variétés clones).

Remarque:

Avec l'hypothèse de dominance, lorsque le caractère sélectionné est gouverné par un grand nombre de loci, l'hétérosis est aussi infixable à l'échelle du sélectionneur.

En conclusion, **la relation entre la dominance et le système de reproduction est immédiate**. L'allogamie permettant le développement du *fardeau génétique*, la fraction d'hétérosis fixable peut être importante pour ce système de reproduction. A l'inverse, le fardeau génétique étant quasiment inexistant chez les espèces autogames (les allèles récessifs défavorables ont été éliminés ou n'ont pas pu se développer sous l'action de la sélection naturelle), ces dernières sont moins sujettes à la *dépression de consanguinité*, ce qui n'exclut pas l'hétérosis dû à la superdominance. La seule possibilité de fixer la superdominance réside dans la duplication du génome diploïde: c'est l'*allopolyploïdie*. Le blé, avec ses trois génomes élémentaires juxtaposés, est un représentant typique de ce processus (le non-appariement à la méiose des génomes répétés confère un comportement diploïde aux espèces allopolyploïdes, contrairement aux espèces autopolyploïdes).

2 Modèle à plusieurs loci

2.1 En l'absence d'épistasie

En l'absence d'épistasie (c'est-à-dire en l'absence d'interaction entre gènes non homologues), la généralisation du modèle précédent au cas d'un caractère quantitatif contrôlé par plusieurs loci indépendants du point de vue de leur action ne présente aucune difficulté. Il suffit simplement de tenir compte de l'additivité des effets des gènes aux différents loci, ce qui se traduit dans les formules par une sommation supplémentaire portant sur l'indice n (identificateur des loci):

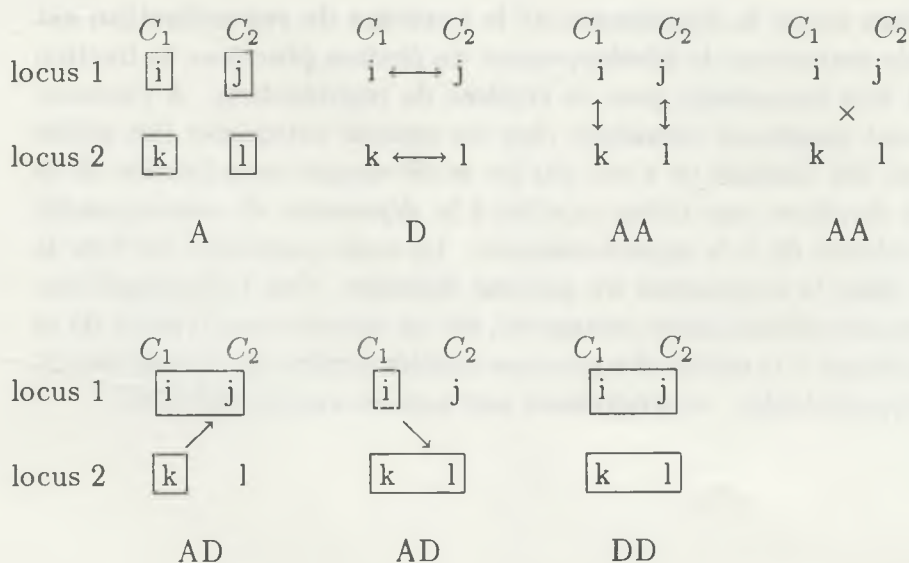
$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 \quad \text{avec} \quad \sigma_A^2 = 2 \sum_n \left(\sum_i p_{i,n} \alpha_{i,n}^2 \right) \quad \text{et} \quad \sigma_D^2 = \sum_n \left(\sum_{ij} p_{i,n} p_{j,n} \beta_{ij,n}^2 \right)$$

2.2 En présence d'épistasie et d'équilibre de liaison

En présence d'épistasie (c'est-à-dire en présence d'interaction entre gènes non homologues) et d'équilibre de liaison (association au hasard des loci dans la population, notion distincte de celle de linkage), la définition des effets procède de la même démarche.

Considérons le cas de deux loci ($n=2$):

Il est alors possible de définir des interactions (cf. ci-dessous) entre deux gènes non homologues (épistasie additive x additive, notée AA), entre deux gènes homologues et un non homologue (épistasie additive x dominance, notée AD) et entre deux gènes homologues et deux autres gènes homologues (épistasie dominance x dominance, notée DD).



Les deux chromosomes sont symbolisés par C_1 et C_2 .

Le modèle de décomposition de la valeur génotypique s'écrit:

$$g_{ijkl} = \begin{array}{ll} \mu + \alpha_i + \alpha_j + \alpha_k + \alpha_l & (A) \\ + \beta_{ij} + \beta_{kl} & (D) \\ + (\alpha\alpha)_{ik} + (\alpha\alpha)_{il} + (\alpha\alpha)_{jk} + (\alpha\alpha)_{jl} & (AA) \\ + (\alpha\beta)_{ikl} + (\alpha\beta)_{jkl} + (\alpha\beta)_{kij} + (\alpha\beta)_{lij} & (AD) \\ + (\beta\beta)_{ijkl} & (DD) \end{array} \left. \vphantom{\begin{array}{l} (A) \\ (D) \\ (AA) \\ (AD) \\ (DD) \end{array}} \right\} \text{épistasie}$$

$$\begin{aligned} \mu &= g_{....} \\ \alpha_i &= g_{i...} - \mu = \sum_{j,k,l} p_j p_k p_l g_{ijkl} - \mu \\ \beta_{ij} &= g_{ij..} - \mu - \alpha_i - \alpha_j = \sum_{k,l} p_k p_l g_{ijkl} - \mu - \alpha_i - \alpha_j \\ (\alpha\alpha)_{ik} &= g_{i.k.} - \mu - \alpha_i - \alpha_k \\ (\alpha\beta)_{ijk} &= g_{ijk.} - \mu - \alpha_i - \alpha_j - \alpha_k - \beta_{ij} - (\alpha\alpha)_{ik} - (\alpha\alpha)_{jk} \end{aligned}$$

On peut, par ailleurs vérifier que:

$$\sum_i p_i \alpha_i = \sum_{ij} p_i p_j \beta_{ij} = \sum_{ik} p_i p_k (\alpha\alpha)_{ik} = \sum_{ijk} p_i p_j p_k (\alpha\beta)_{ijk} = \sum_{ijkl} p_i p_j p_k p_l (\beta\beta)_{ijkl} = 0$$

Du point de vue de l'analyse de variance, on est en présence d'un caractère déterminé par 4 facteurs (père et mère s'exprimant à deux loci différents) et les (β) , $(\alpha\alpha)$, $(\alpha\beta)$ et $(\beta\beta)$ reflètent respectivement les interactions de 1^{er}, 1^{er}, 2^{eme} et 3^{eme} ordre.

Si l'on choisit un individu au hasard dans la population, la variable aléatoire qui mesure sa valeur génotypique s'écrit donc:

$$G = \mu + A + D + AA + AD + DD$$

A: variable aléatoire qui prend pour valeur les $\alpha_i + \alpha_j + \alpha_k + \alpha_l$ (valeur génétique additive),

D: variable aléatoire qui prend pour valeur les $\beta_{ij} + \beta_{kl}$ (valeur génétique de dominance),

AA: variable aléatoire qui prend pour valeur les $(\alpha\alpha)_{ik} + (\alpha\alpha)_{jl} + (\alpha\alpha)_{jk} + (\alpha\alpha)_{il}$ (valeur génétique additive \times additive),

AD: variable aléatoire qui prend pour valeur les $(\alpha\beta)_{ikl} + (\alpha\beta)_{jkl} + (\alpha\beta)_{kij} + (\alpha\beta)_{lij}$ (valeur génétique additive \times dominance),

DD: variable aléatoire qui prend pour valeur les $(\beta\beta)_{ijkl}$ (valeur génétique dominance \times dominance).

On a:

$$E(A) = E(D) = E((AA)) = E((AD)) = E((DD)) = 0$$

Une variance est définie pour chaque type d'effet:

$$\begin{aligned} \sigma_{AA}^2 &= 4E((AA)^2) \\ \sigma_{AD}^2 &= 2E((AD)^2) + 2E((DA)^2) \end{aligned}$$

$$\sigma_{DD}^2 = E((DD)^2)$$

L'hypothèse d'équilibre de liaison permettant d'annuler les covariances entre loci, la variance génétique totale s'écrit:

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2$$

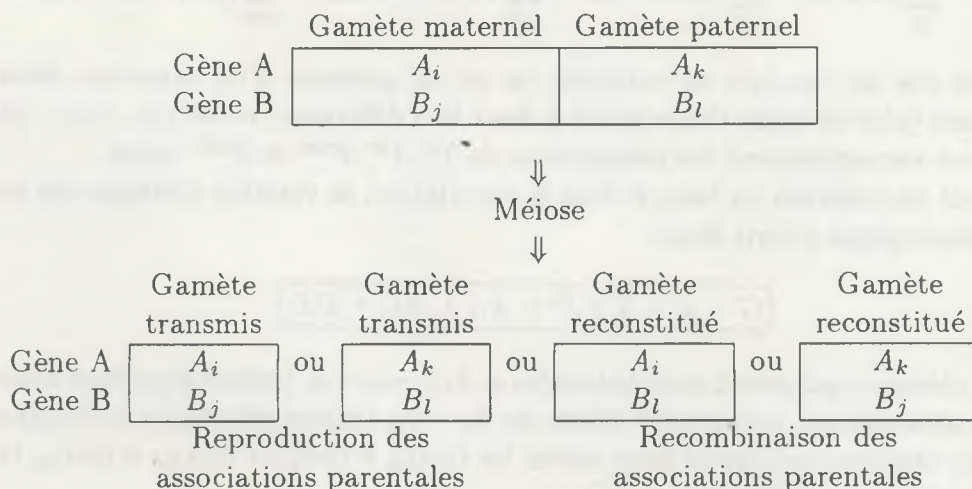
On notera I la somme des phénomènes d'épistasie (AA) , (AD) et (DD) .

L'extension du modèle à un nombre supérieur à deux loci ne pose pas de problème autre que celui d'interpréter les nombreux paramètres qui en découlent (interactions géniques d'ordre supérieur à deux AAA, AAD, ADD, DDD,...).

2.3 Le déséquilibre de liaison

En ne considérant que deux loci, le *déséquilibre de liaison* traduit l'écart entre la fréquence de chaque type de gamète et l'espérance de cette fréquence si les gènes étaient associés au hasard.

Le schéma ci-dessous figure la transmission et la recombinaison des allèles:



Si les deux loci sont indépendants, les quatre types de gamètes seront équiprobables, par contre s'il n'y a pas d'indépendance, la transmission aura tendance à favoriser les associations parentales.

Désignons par c la proportion de recombinaisons:

- si $c = 0$ → il n'y a pas de recombinaison (linkage total=liaison physique des loci sur le gamète),
- si $c = 1/2$ → il y a autant de recombinaisons que de reproductions des associations parentales (loci indépendants).

Notons P_{AB} la fréquence d'une combinaison AB dans la population parentale et P'_{AB} la fréquence de cette même combinaison à la génération suivante, on a :

$$P'_{AB} = P_{AB}(1 - c) + c p_A p_B$$

où p_A et p_B sont les fréquences des allèles A et B .

$\Delta = P_{AB} - p_A p_B$ est le *Déséquilibre de liaison*.

Dans le cas biallélique, notons par l'indice 1 les allèles favorables aux deux loci (A_1 et B_1) et par l'indice 2 les allèles défavorables (A_2 et B_2). On pourra écrire :

$$\Delta = P_{A_1 B_1} - p_{A_1} p_{B_1} = P_{A_2 B_2} - p_{A_2} p_{B_2} = -(P_{A_1 B_2} - p_{A_1} p_{B_2}) = -(P_{A_2 B_1} - p_{A_2} p_{B_1}) = P_{A_1 B_1} P_{A_2 B_2} - P_{A_1 B_2} P_{A_2 B_1}$$

Le signe de Δ est arbitraire. Par convention :

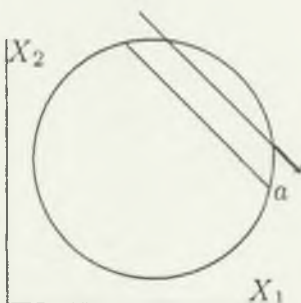
$$\begin{cases} \Delta > 0 & \text{lorsqu'il y a excès de gamètes portant des allèles favorables (couplage),} \\ \Delta < 0 & \text{lorsqu'il y a excès de gamètes portant des allèles défavorables (répulsion),} \\ \Delta = 0 & \text{dans la situation d'équilibre.} \end{cases}$$

Il serait plus juste de parler de *déséquilibre gamétique* ou de *déséquilibre d'association* puisque la liaison physique n'intervient pas. Ce déséquilibre est souvent le reflet d'un phénomène de coadaptation génique (arrangement particulier d'allèles dont la transmission confère un avantage sélectif).

Lorsqu'il y a *équilibre de liaison* la covariance entre les effets des allèles à deux loci différents d'un même gamète est nulle.

Pour mieux visualiser l'**impact de la sélection sur l'équilibre de liaison**, on considère un caractère Y déterminé par la somme de deux variables X_1 et X_2 indépendantes. La distribution de Y est, dans l'exemple schématisée par un cercle, les valeurs de Y sont constantes sur une parallèle à la deuxième bissectrice.

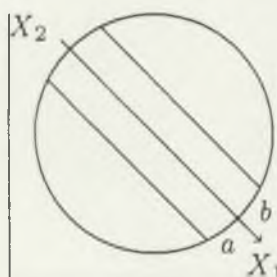
Schématiquement, trois types de sélection peuvent tronquer la distribution de Y :



Sélection directionnelle

$$Y > a$$

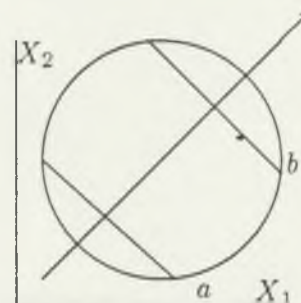
$$\text{cov}(X_1, X_2) < 0$$



Sélection pour un optimum

$$a < Y < b$$

$$\text{cov}(X_1, X_2) < 0$$



Sélection disruptive

$$Y < a \text{ et } Y > b$$

$$\text{cov}(X_1, X_2) > 0$$

En considérant les X comme les effets additifs des gènes sur le caractère, et sachant l'équivalence entre covariance et déséquilibre de liaison dans le cas additif, on constate qu'une sélection directionnelle ou pour un optimum générera des $\Delta < 0$, tandis qu'une sélection disruptive générera des $\Delta > 0$ entre gènes à des loci différents.

Autres facteurs affectant l'équilibre de liaison:

- *le système de reproduction*: par accouplements non-aléatoires ou structuration des populations,
- *la dérive*: l'échantillonnage entraîne des variations aléatoires de la valeur de Δ ,
- *les phénomènes d'entraînement* ("auto-stop" = "hitch-hiking"): ils concernent les variations de fréquence en un locus, neutre ou sélectionné, en fonction de son déséquilibre de liaison avec un locus sélectionné (Δ initial doit être $\neq 0$),
- *le nombre de loci*: on met en évidence des déséquilibres de liaison d'autant plus importants que le nombre de loci pris en compte dans le modèle est élevé.

Remarque:

La liaison génétique en elle-même n'engendre pas de déséquilibre de liaison. Si un déséquilibre Δ_{ij} existe ou est créé à la génération (t) entre deux loci i et j , sa relaxation à la génération suivante en l'absence des facteurs l'ayant généré sera d'autant plus lente que le taux de recombinaison c_{ij} entre ces loci sera plus faible:

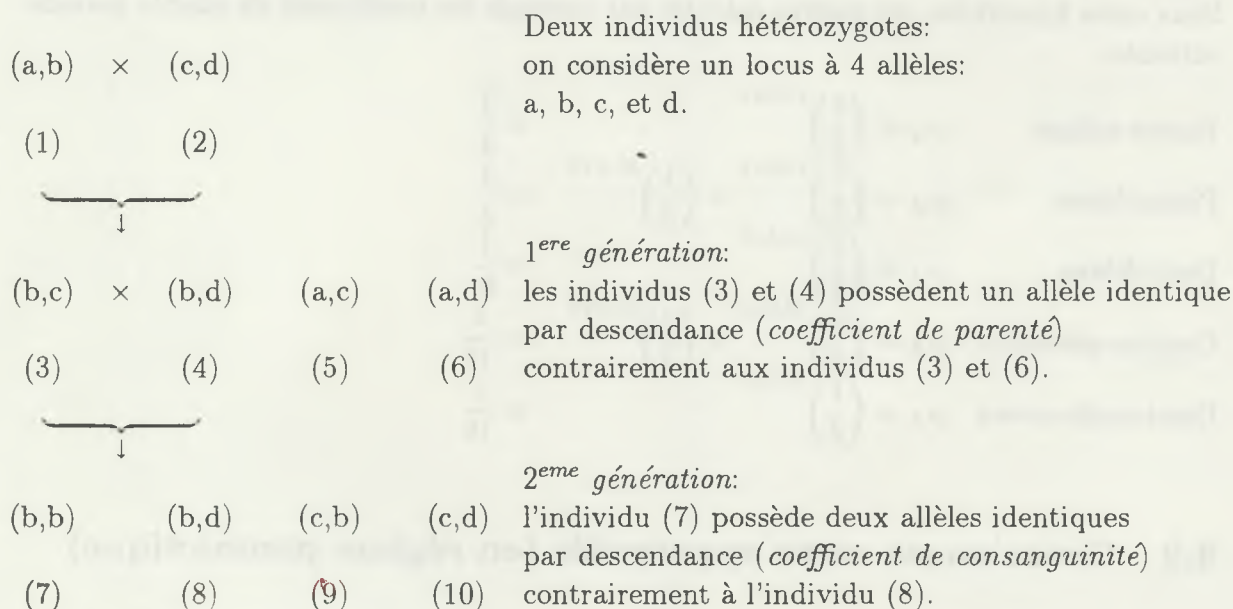
$$\Delta_{ij}(t+1) = (1 - c_{ij})\Delta_{ij}(t)$$

3 Notions d'apparentement

Identité des gènes: Deux gènes sont identiques s'ils dérivent sans mutation (par descendance) d'un même gène ancêtre.

On appelle *coefficient d'identité* la probabilité pour qu'à un locus les deux allèles soient identiques par descendance. Lorsque cette probabilité est définie pour un seul individu, il s'agit du *coefficient de consanguinité*, lorsqu'elle est définie pour deux individus, on parle de *coefficient de simple parenté*.

Le schéma suivant figure différentes relations d'identité à partir du croisement de deux individus non apparentés ((1) et (2)):



3.1 Coefficient de simple parenté

Le calcul du coefficient de simple parenté φ_A fait intervenir la généalogie des deux individus et le coefficient de consanguinité de leur(s) ancêtre(s) commun(s).

Supposons que X et Y n'aient qu'un ancêtre commun : Z. Un gène pris au hasard chez X ne peut être identique à un gène pris au hasard chez Y que s'ils sont tous deux:

- soit la réplique du même gène de l'ancêtre commun Z,
- soit la réplique des deux gènes de Z, ceux-ci étant identiques en raison de la consanguinité de Z (F_Z étant le coefficient de consanguinité de Z).

Si Z est un ancêtre de degré n de X et de degré p de Y, on a:

$$\varphi_A = \left(\frac{1}{2}\right)^n \left(\frac{1}{2}\right)^p \left[\frac{1}{2}(1 - F_Z) + F_Z \right] = \left(\frac{1}{2}\right)^{n+p+1} (1 + F_Z)$$

S'il y a plusieurs ancêtres i communs sans lien de parenté entre eux:

$$\varphi_A = \sum_i \left(\frac{1}{2}\right)^{n_i+p_i+1} (1 + F_Z)$$

On suppose généralement que les ancêtres communs sont non-consanguins et non-apparentés entre eux.

Sous cette hypohèse, on pourra calculer par exemple les coefficients de simple parenté suivants:

Parent-enfant	$\varphi_A = \left(\frac{1}{2}\right)^{1+0+1}$	$= \frac{1}{4}$
Pleins-frères	$\varphi_A = \left(\frac{1}{2}\right)^{1+1+1} + \left(\frac{1}{2}\right)^{1+1+1}$	$= \frac{1}{4}$
Demi-frères	$\varphi_A = \left(\frac{1}{2}\right)^{1+1+1}$	$= \frac{1}{8}$
Cousins-germains	$\varphi_A = \left(\frac{1}{2}\right)^{2+2+1} + \left(\frac{1}{2}\right)^{2+2+1}$	$= \frac{1}{16}$
Demi-oncle-neveu	$\varphi_A = \left(\frac{1}{2}\right)^{1+2+1}$	$= \frac{1}{16}$

3.2 Covariances entre apparentés (en régime panmictique)

Grâce à la définition par Malécot (1948) du concept d'identité entre gènes, on peut écrire la covariance entre deux individus X et Y ayant un degré de parenté donné en fonction de deux paramètres: le *coefficient de simple parenté* et le *coefficient de double parenté*. Dans une population panmictique d'effectif limité, la probabilité pour que deux allèles portés par deux individus différents dérivent par réplication d'un même ancêtre commun définit le *coefficient simple de parenté* φ_A . La probabilité de tirer, dans une famille d'apparentés, un individu portant deux allèles identiques à ceux d'un autre individu de la même famille définit le *coefficient de double parenté* φ_D .

Soit deux individus X et Y non consanguins et entre lesquels il existe un lien de parenté.

On note:

$a_f a_m$ et $a'_f a'_m$ leurs génotypes, p_i la fréquence de a_i et α_i son effet moyen, les indices f et m indiquant respectivement l'origine maternelle et paternelle des allèles,

les valeurs génotypiques G et G' de X et Y s'écrivent:

$$G = \mu + A_f + A_m + D \quad \text{et} \quad G' = \mu + A'_f + A'_m + D'$$

La covariance génétique entre ces deux individus apparentés s'écrit:

$$\begin{aligned} \text{cov}(G, G') &= \text{cov}(A_f, A'_f) + \text{cov}(A_f, A'_m) + \text{cov}(A_m, A'_f) + \text{cov}(A_m, A'_m) \\ &\quad + \text{cov}(A_f, D') + \text{cov}(A_m, D') + \text{cov}(D, A'_f) + \text{cov}(D, A'_m) \\ &\quad + \text{cov}(D, D') \end{aligned}$$

* Calcul des termes du type $E(A_x A_y)$ avec $x = f$ ou m et $y = f$ ou m .

$$E(A_x A_y) = E(A_x A_y / a_x \equiv a_y) \cdot P(a_x \equiv a_y) + E(A_x A_y / a_x \neq a_y) \cdot P(a_x \neq a_y)$$

$a_x \equiv a_y$ signifie que a_x et a_y sont identiques par descendance.

$$E(A_x A_y) = \sum_i p_i \alpha_i^2 P(a_x \equiv a_y) + \sum_{i \neq j} p_i p_j \alpha_i \alpha_j P(a_x \neq a_y)$$

$$\sum_{i \neq j} p_i p_j \alpha_i \alpha_j = 0 \quad \text{et} \quad \sum_i p_i \alpha_i^2 = \frac{1}{2} \text{var}(A_f + A_m) = \frac{1}{2} \sigma_A^2$$

d'où:

$$\begin{aligned} &\text{cov}(A_f, A'_f) + \text{cov}(A_f, A'_m) + \text{cov}(A_m, A'_f) + \text{cov}(A_m, A'_m) \\ &= \frac{\sigma_A^2}{2} [P(a_f \equiv a'_f) + P(a_f \equiv a'_m) + P(a_m \equiv a'_f) + P(a_m \equiv a'_m)] \\ &= \frac{\sigma_A^2}{2} [4\varphi_A] = 2\sigma_A^2 \varphi_A \quad \text{avec } \varphi_A \text{ le coefficient de simple parenté de } X \text{ et } Y. \end{aligned}$$

* Calcul des termes du type $E(A_x D_{yz})$ avec $a_x \equiv a_y$ ou $a_x \equiv a_z$.

$$E(A_x D_{yz}) = \sum_{ij} \alpha_i d_{ij} p_i p_j P(a_x \equiv a_y) + \sum_{ij} \alpha_i d_{ij} p_i p_j P(a_x \equiv a_z)$$

$$\text{avec } \sum_{ij} \alpha_i d_{ij} p_i p_j = \sum_i \alpha_i p_i \left[\sum_j p_j d_{ij} \right] = 0$$

* Calcul de $E(D_{fm} D'_{fm})$ avec $\begin{pmatrix} a_f \equiv a'_f \\ a_m \equiv a'_m \end{pmatrix}$ ou $\begin{pmatrix} a_f \equiv a'_m \\ a_m \equiv a'_f \end{pmatrix}$.

$$\begin{aligned} E(D_{fm} D'_{fm}) &= \sum_{ij} p_i p_j d_{ij}^2 [P(a_f \equiv a'_f, a_m \equiv a'_m) + P(a_f \equiv a'_m, a_m \equiv a'_f)] \\ &= \sigma_D^2 \varphi_D^2 \quad \text{avec } \varphi_D \text{ le coefficient de double parenté de } X \text{ et } Y. \end{aligned}$$

donc:

$$\boxed{\text{cov}(G, G') = 2\varphi_A \sigma_A^2 + \varphi_D \sigma_D^2}$$

La figure ci-dessous illustre les parentés demi-frères et pleins-frères (Gallais, 1990).

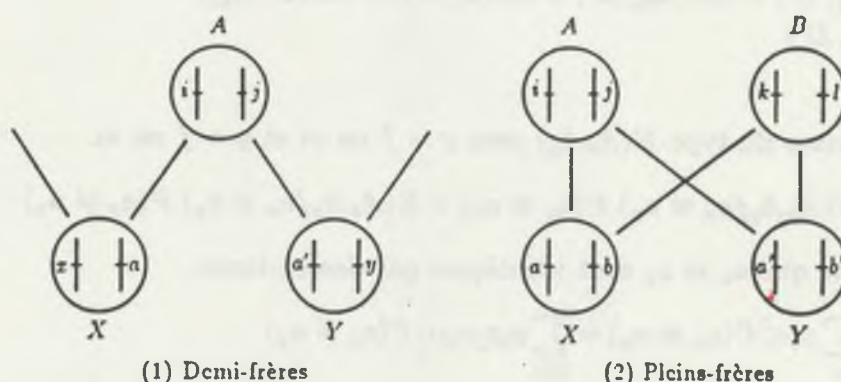


Figure 2.19 : Illustration des parentés demi-frères et pleins-frères. Au locus considéré a et a' représentent les gènes venant de A (i ou j) et b et b' les gènes venant de B (k ou l); pour les demi-frères x et y sont 2 gènes quelconques issus d'individus non apparentés à A . Pour les demi-frères $\varphi_A(HS) = (\text{Prob. de tirer } a \text{ et } a') \times [\text{Prob. } (a \equiv a') \text{ sachant que } a \text{ et } a' \text{ ont été tirés}]$. Or, $\text{Prob. } (a \equiv a')/a, a' = \text{Prob. de tirer 2 gènes identiques dans } A$; c'est le coefficient de parenté d'un individu avec lui-même. D'où $\varphi_A(HS) = \frac{1}{8}$.

Estimation de la covariance parent-enfant P-O (en anglais: Parent-Offspring).

La panmixie permet d'écrire : $\mu_{\text{parent}} = \mu_{\text{enfant}} = \mu$

$$\text{cov}(PO) = E[(G_{\text{parent}} - \mu)(G_{\text{enfant}} - \mu)] = \frac{1}{2}\sigma_A^2$$

Estimation de la covariance demi-frères HS (en anglais: Half-Sib).

Soit $i(pa)$ et $j(pa)$ deux enfants demi-frères d'un même parent pa :

$$\text{cov}(HS) = E_{pa}[(G_{i(pa)} - \mu)(G_{j(pa)} - \mu)] = \frac{1}{4}\sigma_A^2$$

Quelques exemples de covariances entre apparentés déduites des coefficients de simple et de double parenté sont présentés dans le tableau ci-dessous :

Parenté	φ_A	φ_D	$cov(X, Y)$
Parent-enfant (<i>PO</i>)	$\frac{1}{4}$	0	$\frac{1}{2}\sigma_A^2$
Pleins-frères (<i>FS</i>)	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_D^2$
Demi-frères (<i>HS</i>)	$\frac{1}{8}$	0	$\frac{1}{4}\sigma_A^2$
Cousins-germains	$\frac{1}{16}$	0	$\frac{1}{8}\sigma_A^2$
Demi-oncle-neveu (<i>HUN*</i>)	$\frac{1}{16}$	0	$\frac{1}{8}\sigma_A^2$
1 individu avec lui-même	$\frac{1}{2}(**)$	1	$\sigma_A^2 + \sigma_D^2$

* en anglais: Half-Uncle-Nephew.

(**) Le coefficient de simple parenté d'un individu avec lui-même n'est égal à 1 que si cet individu est homozygote.

La connaissance de certaines covariances entre apparentés permettra ainsi de déduire les variances d'additivité et de dominance et donc la variance génétique totale.

Dans le cas de prédiction linéaire, estimer la covariance entre apparentés permettra de prévoir la valeur des descendants d'une plante.

4 Effets des systèmes de reproduction

Au cours de la sélection, les systèmes de reproduction vont agir sur la structure génotypique de la population. Nous allons étudier l'impact des différents systèmes de reproduction sur la moyenne de la population.

Le tableau suivant récapitule les **taux d'allogamie** (pourcentages de fécondations croisées) des principales espèces étudiées au CIRAD.

Cacaoyer		de 75% à 95%
Caféier	. arabica	≈ 10%
	. robusta	> 95%
Cocotier	. grand ouest africain	de 70% à 100%
	. nain	de 0% à 50%
Coton	↘ lorsque le niveau d'insecticide ↗ dépend également des densités	de 7% à 35%
Eucalyptus	. mesure sur graine	de 70% à 90%
	. stade jeune plant	≈ 90%
Hévéa		de 75% à 95%
Palmier à huile		> 95%
Riz	dépend des conditions de culture	de 2% à 10%
Sorgho		de 6% à 70%

Remarque:

Tandis que le taux d'autofécondation chez les arbres entomogames (fécondation par les insectes) avoisine 10%, il n'est plus que de 5% chez les arbres anémogames (fécondation par le vent) tels que les gymnospermes (pins).

4.1 La consanguinité

L'autofécondation permet de diminuer le taux d'hétérozygotie à chaque génération puisque par autofécondation les homozygotes donnent des homozygotes et les hétérozygotes donnent seulement 50% d'hétérozygotes. Ainsi, sous l'hypothèse d'indépendance des loci, à chaque génération d'autofécondation le taux d'hétérozygotie est divisé par 2.

Considérons une population panmictique, c'est-à-dire où il n'existe pas d'individus consanguins. Sa valeur moyenne est égale à μ , qui est la moyenne entre la population homozygote de moyenne μ_L et la population hétérozygote.

$$\mu = \sum_{i,j} p_i p_j g_{ij} \quad \sum_{i \neq j} p_i p_j g_{ij} + \sum_i p_i^2 g_{ii}$$

$$\mu_L = \sum_i p_i^2 g_{ii}$$

Si la population est infinie et s'il y a panmixie, on retrouve à la génération suivante une population identique, mais si par autofécondation ou par croisements dirigés on constitue

dans une proportion F une sous-population d'individus consanguins (donc homozygotes et de valeur moyenne μ_L), on obtient une population consanguine de moyenne μ_F :

$$\mu_F = (1 - F)\mu + F\mu_L = \mu + F(\mu_L - \mu)$$

F est le *coefficient de consanguinité* de la population.

$(\mu_L - \mu) = D_0$ est la *dépression de consanguinité maximale*. En l'absence d'épistasie, l'expression de la valeur de μ_F est vérifiée quel que soit le nombre de loci concernés.

Le modèle de décomposition de la valeur génotypique des homozygotes s'écrit:

$$g_{ii} = \mu + 2\alpha_i + \beta_{ii}$$

g_{ii} est la valeur prise par la variable aléatoire $G = \mu + A + D_0$

D_0 étant l'interaction entre allèles homologues (dominance) restreinte aux β_{ii} ($E(D_0) = d_0 \neq 0$)

Donc la moyenne μ_L s'écrit:

$$\mu_L = \sum_i p_i^2 g_{ii} = \mu + \sum_i p_i^2 \beta_{ii} \quad (\text{avec } E(A) = \sum_i p_i \alpha_i = 0)$$

d'où: $\mu_F = \mu + F \cdot E(D_0)$

Pour un ensemble de loci, en l'absence d'épistasie, il suffit de réaliser une sommation portant sur l'indice n (identificateur des loci):

$$\mu_F = \mu + F \sum_n E(D_0) = \mu + F \cdot \sum_n d_0$$

On retrouve bien $\mu_F = \mu_L$ lorsque $F = 1$.

Ainsi, en l'absence d'épistasie et pour une valeur de F donnée, la consanguinité entraînera toujours une dépression de consanguinité, quel que soit le degré de dominance, si celle-ci est favorable ($d > 0$), comme l'indique la figure suivante (Gallais, 1990).

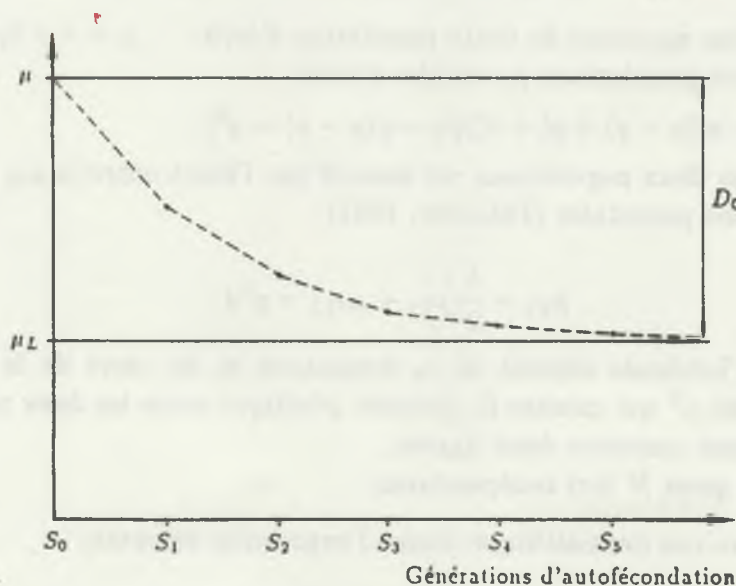


Figure 2.21 : Effet attendu de l'autofécondation sur la vigueur d'une population. (en l'absence d'épistasie et sans sélection); μ : moyenne de la population panmictique; μ_L : moyenne de la population à l'état homozygote; D_0 : dépression maximale de consanguinité.

4.2 Hybridation entre populations

On considère deux populations bialléliques P_1 et P_2 , dont les fréquences respectives pour les deux allèles sont p_1, q_1 et p_2, q_2 .

Soit F_1 , la population constituée par les descendants du croisement entre les deux populations P_1 et P_2 , en utilisant les paramètres du modèle génétique, on peut écrire:

$$\begin{array}{ccccc}
 & P_1 & & P_2 & \\
 A_1 & A_2 & \times & A_1 & A_2 \\
 p_1 & q_1 & & p_2 & q_2 \\
 \hline
 A_1A_1 & A_1A_2 & & A_2A_1 & A_2A_2 \\
 p_1p_2 & p_1q_2 & & p_2q_1 & q_1q_2
 \end{array}$$

donc $\mu_{F_1} = p_1p_2g_{11} + (p_1q_2 + p_2q_1)g_{12} + q_1q_2g_{22}$

avec $g_{11} = a + c$ $g_{22} = c - a$ $g_{12} = d + c$

$$\begin{aligned}
 \mu_{F_1} &= p_1p_2(a + c) + (p_1q_2 + p_2q_1)(d + c) + q_1q_2(c - a) \\
 &= c(p_1p_2 + q_1q_2 + p_1q_2 + p_2q_1) + a(p_1p_2 - q_1q_2) + d(p_1q_2 + p_2q_1) \\
 &= c + a(p_1p_2 - q_1q_2) + d(p_2q_1 + q_2p_1)
 \end{aligned}$$

On pose $p_2 = p_1 + y$ et $p_1 = p$:

$$\mu_{F_1} = c + a[(p - q) + y] + d[2pq - y(p - q)]$$

Sachant que la valeur moyenne de toute population s'écrit: $\mu = c + 2p q d + (p - q)a$
la moyenne des deux populations parentales s'écrit:

$$\frac{1}{2}(\mu_{P_1} + \mu_{P_2}) = c + a[(p - q) + y] + d[2pq - y(p - q) - y^2]$$

L'hétérosis entre ces deux populations est mesuré par l'écart entre la μ_{F_1} et la moyenne des deux populations parentales (Falconer, 1961):

$$\mu_{F_1} - \frac{1}{2}(\mu_{P_1} + \mu_{P_2}) = y^2 d$$

On constate que l'hétérosis dépend de la dominance et du carré de la différence des fréquences alléliques, y^2 qui mesure la *distance génétique* entre les deux populations.

Lorsque le croisement concerne deux lignées,

$y = 1$ et $d = \sum_n d_n$ pour N loci indépendants.

La généralisation au cas multiallélique donne l'expression suivante:

$$\mu_{F_1} - \frac{1}{2}(\mu_{P_1} + \mu_{P_2}) = \frac{1}{2} \sum_{i,j} (p_{2i} - p_{1i})(p_{2j} - p_{1j}) d_{ij}$$

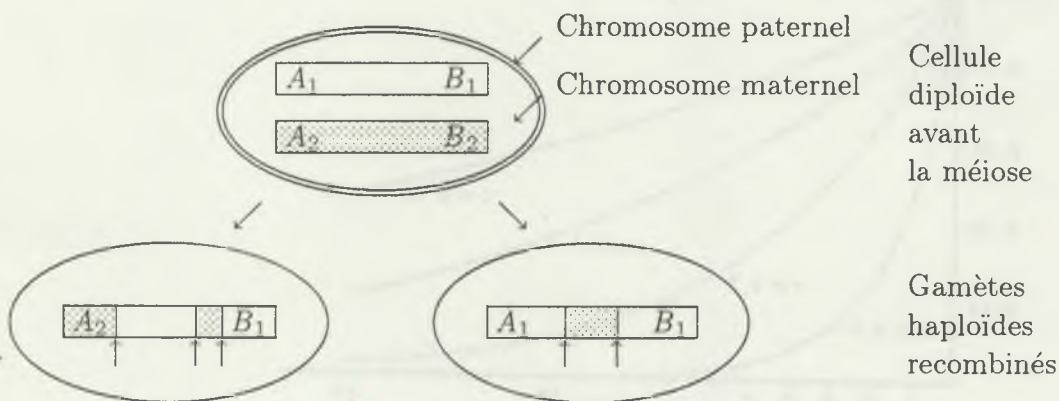
avec p_{1i} et p_{2i} les fréquences respectives de l'allèle i dans la population 1 et dans la population 2.

On voit qu'il peut exister de l'hétérosis entre deux populations sans qu'il y ait consanguinité, ainsi *l'hétérosis n'est plus le corollaire de la dépression de consanguinité*.

4.3 Effet de la panmixie

La succession de plusieurs générations d'intercroisement dans une population aura pour effet d'accumuler le nombre de méioses et donc de recombinaisons (crossing-over). La panmixie étant un régime de reproduction où les gamètes mâles et femelles se rencontrent de manière totalement aléatoire vis-à-vis du ou des loci considérés, une population panmictique tendra vers l'équilibre de liaison.

Le schéma suivant représente (selon Bulmer) les effets de la recombinaison à partir du croisement de deux génotypes $A_1A_1B_1B_1$ et $A_2A_2B_2B_2$ (deux loci bialléliques). Le gamète A_1B_1 peut reproduire une association gamétique de la génération précédente ou bien résulter d'une recombinaison.



Les flèches verticales indiquent les emplacements des crossing-over.

Soit c la probabilité de recombinaison entre deux gamètes, p_A la probabilité qu'un gamète porte l'allèle A_1 et p_B la probabilité qu'il porte l'allèle B_1 , on retrouve la fréquence P_{AB_m} des gamètes A_1B_1 à la génération m qui s'écrit:

$$P(AB)_m = (1 - c)P_{AB_{m-1}} + cp_Ap_B$$

La probabilité G_m de tirer chez un individu, à la génération de panmixie m , deux allèles non homologues issus d'une même association gamétique ancêtre, est le *coefficient de lien* entre ces deux allèles (G est analogue au *coefficient de consanguinité* F):

$$G_m = (1 - c)^{m-1}G_1$$

Remarque:

Alors que les coefficients d'identité concernent les allèles homologues (à un locus), le coefficient de lien concerne des allèles non-homologues (à plusieurs loci).

La moyenne μ_m à la génération m de panmixie s'écrit (Gallais, 1974):

$$\mu_m = \mu + 4G_mE(AA)$$

avec μ la moyenne de la population panmictique à l'équilibre de liaison ($m = \infty$) et $E(AA)$ l'espérance des effets d'épistasie additive \times additive.

L'évolution du coefficient de lien au cours des générations d'intercroisement est présentée sur la figure suivante, en fonction du taux de recombinaisons c .

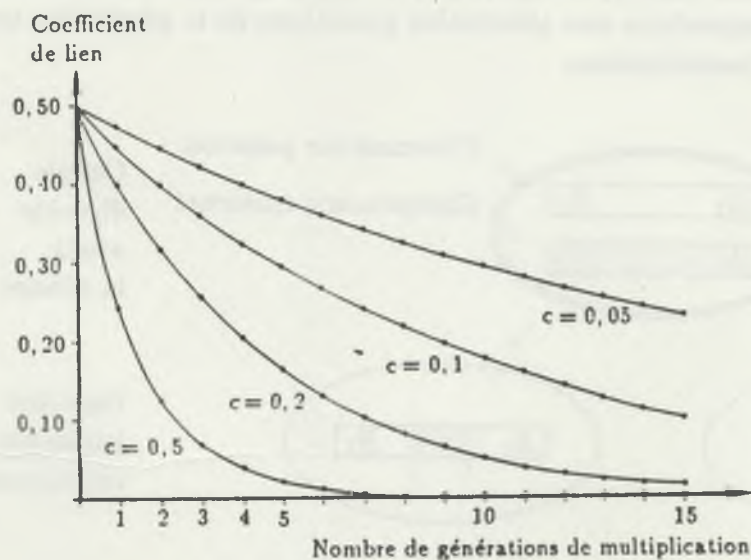


Figure 2.34 : Évolution du coefficient de lien au cours des générations de multiplication en panmixie (d'après BARADAT, 1982).

5 Génotype et Phénotype

Lorsque l'on prend un individu i au hasard dans la population, sa valeur génotypique g_i est la valeur prise par une variable aléatoire G , puisque l'individu est tiré au sort.

Or les mesures effectuées sur la plante entière observée sur le terrain donnent accès au phénotype de la plante. Ce phénotype correspond à "l'ensemble des caractères visibles résultant de l'expression du génotype dans un milieu donné" (Sournia *et al*, 1991). Le génotype n'est alors qu'un concept défini par l'ensemble des expressions possibles d'une combinaison particulière de gènes. Ainsi, le résultat d'une mesure sur un individu particulier est la valeur prise par une variable aléatoire P qui s'écrit:

$$P = G + E = \mu + A + D + I + E$$

où I recouvre toutes les interactions d'épistasie.

E est une variable aléatoire d'espérance supposée nulle qui rend compte des variations environnementales (effet milieu).

Remarque sur E :

Les effets milieux sont de plusieurs types. Certains sont plus ou moins systématiques (variations saisonnières) et parfois contrôlables (conditions de culture), d'autres sont le résultat de facteurs difficilement contrôlables et de microphénomènes pouvant être propres à chaque individu.

Remarque sur E et G :

On suppose généralement qu'il y a indépendance entre l'effet génotype et l'effet milieu, c'est-à-dire qu'il n'y a pas d'interaction génotype \times milieu ($cov(GE) = 0$). Dans le cas contraire les conclusions à venir seraient remises en question.

Si, de plus, on suppose qu'il n'y a pas d'épistasie, le modèle simplifié s'écrit:

$$P = \mu + A + D + E$$

avec $E(A) = E(D) = E(E) = 0$

La variance phénotypique s'écrit:

$$\sigma_P^2 = \sigma_A^2 + \sigma_D^2 + \sigma_E^2$$

Les variables aléatoires A , D et E sont supposées, le plus souvent, de loi Normale. La valeur génétique peut alors être prévue en connaissant la valeur phénotypique par prédiction linéaire.

5.1 Héritabilités

Héritabilité au sens large: h_{SL}^2

L'héritabilité au sens large donne une indication sur la limite de l'amélioration possible du matériel étudié (toute chose étant égale par ailleurs). C'est le degré de confiance dans la prédiction de la valeur génotypique par la valeur phénotypique.

$$h_{SL}^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_E^2}$$

Sachant que $cov(GE) = 0$, $cov(PG) = \sigma_G^2$ donc $h_{SL}^2 = \frac{cov(PG)}{var P}$

L'héritabilité au sens large est le coefficient de régression de G par rapport à P .
Il est clair que diminuer σ_E^2 c'est augmenter h_{SL}^2 .

Héritabilité au sens strict: h_{SS}^2

L'héritabilité au sens strict donne une indication sur les gains que l'on peut espérer en créant des populations améliorées. Elle fait intervenir le système de reproduction et la modélisation de son fonctionnement. C'est le degré de confiance dans la prédiction de la valeur des enfants (en croisement) par la valeur phénotypique des parents.

$$h_{SS}^2 = \frac{\sigma_A^2}{\sigma_G^2 + \sigma_E^2}$$

Deux articles développant le concept d'héritabilité figurent en Annexes.

5.2 Covariances entre caractères

L'amélioration des plantes est le plus souvent multicaractère, d'un point de vue tant économique que biologique. La connaissance des relations entre caractères permet de mieux transformer la plante (dans la mesure où tout caractère lié à un autre caractère apporte une information sur ce dernier) mais aussi de mieux la comprendre dans ses relations avec le milieu.

Sous l'hypothèse d'indépendance entre les génotypes et les milieux, la covariance phénotypique entre deux caractères 1 et 2 se décompose de la manière suivante:

$$cov P_1 P_2 = cov[(G_1 + E_1), (G_2 + E_2)] = cov G_1 G_2 + cov E_1 E_2$$

Dans le cas de nombreuses variables, on peut donc écrire:

$$\Sigma_{PP} = \Sigma_{GG} + \Sigma_{EE}$$

où Σ_{PP} , Σ_{GG} et Σ_{EE} sont respectivement les matrices de variances-covariances phénotypique, génotypique et environnementale.

Si les corrélations génétiques et environnementales sont de signes opposés, une faible corrélation phénotypique entre deux caractères ne signifiera pas nécessairement une faible corrélation génotypique (cf Figure ci-dessous). Les corrélations environnementales dépendent largement de la nature du facteur milieu. De façon générale, toutes ces corrélations sont définies pour une population donnée de génotypes et une population donnée de milieux.

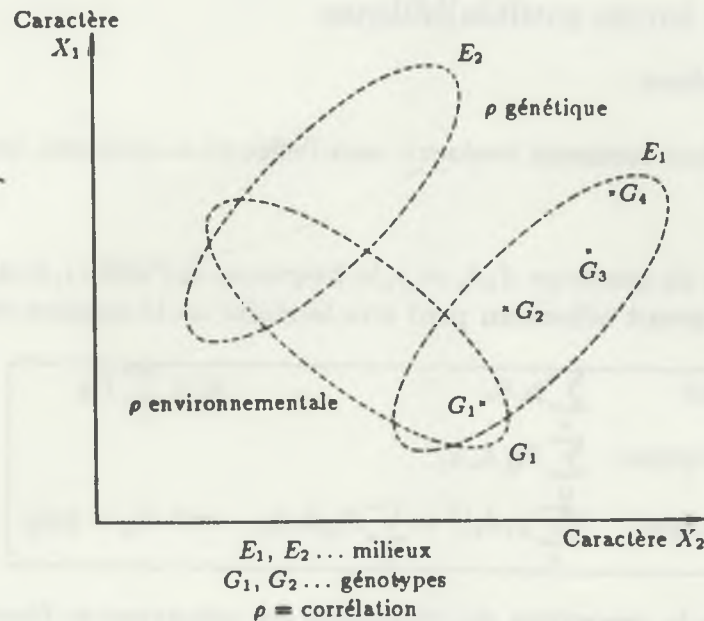


Figure 2.10 : Illustration des corrélations génétiques et des corrélations environnementales entre caractères. Pour bien les distinguer, ces corrélations ont été prises de sens inverse. Ce n'est évidemment pas une situation générale (voir texte).

Signification biologique

Trois phénomènes peuvent être à l'origine de la liaison génétique entre deux caractères:

- . le *linkage* dû à l'effet de la sélection naturelle qui regroupe au sein d'une même unité de ségrégation les gènes dont la combinaison présente un avantage sélectif;
- . la *pléiotropie*, situation dans laquelle une unité de ségrégation agit sur plusieurs caractères, qui est l'aboutissement du processus de sélection naturelle;
- . le *déséquilibre de liaison* dû à l'effet de dérive (variation aléatoire de la fréquence génétique dans des populations d'effectif limité) qui entraîne des associations sans signification génétique.

6 Théorème fondamental de la sélection

6.1 Cas d'un locus multiallélique

Position du problème :

Il s'agit d'évaluer comment évoluent, sous l'effet de la sélection, les paramètres de la population.

Soit P_{ij} la fréquence du génotype $A_i A_j$ et p_i la fréquence de l'allèle i , la **caractérisation de la population avant sélection** peut être formulée de la manière suivante:

Distribution génique	$\sum_i p_i A_i$	$p_i = \sum_j P_{ij}$
Distribution génotypique	$\sum_{ij} P_{ij} A_i A_j$	
En régime panmictique	$(\sum_i p_i A_i)^2 = \sum_{ij} P_{ij} A_i A_j \iff P_{ij} = p_i p_j$	

La sélection modifie la répartition des fréquences des génotypes en fonction de la probabilité de survie s_{ij} attachée à chaque génotype $A_i A_j$.

La probabilité moyenne de survie $S = \sum_{ij} P_{ij} \cdot s_{ij}$

La nouvelle fréquence P'_{ij} , après sélection, du génotype $A_i A_j$ s'écrit comme la probabilité qu'un individu pris au hasard dans la population sélectionnée ait le génotype $A_i A_j$ sachant qu'il a survécu.

$$P'_{ij} = \frac{P_{ij} \cdot s_{ij}}{\sum_{ij} P_{ij} \cdot s_{ij}} = P_{ij} \frac{s_{ij}}{S}$$

Sous l'hypothèse de panmixie : $P'_{ii} = p_i^2 \frac{s_{ii}}{S}$ et $P'_{ij} = 2p_i p_j \frac{s_{ij}}{S}$

Ces expressions montrent qu'il n'est pas nécessaire de connaître les probabilités de survie s_{ij} mais seulement des nombres m_{ij} qui leur sont proportionnels : les *valeurs sélectives*. En cohérence avec la définition de la valeur sélective génotypique, on mesure la valeur sélective de l'allèle A_i par:

$$m_i = \sum_j p_j m_{ij}$$

En l'absence d'effet lié au sexe, la valeur sélective du génotype $A_i A_j$: $m_{ij} = m_{ji}$

paramètres génotypiques	paramètres géniques
$(A_i A_j, P_{ij}, m_{ij})$	(A_i, p_i, m_i)

Soit la valeur sélective moyenne $\bar{m} = \sum_i p_i m_i = \sum_{i,j} P_{ij} m_{ij} = \sum_{i,j} p_i p_j m_{ij}$

la caractérisation de la population après sélection peut alors être formulée de la manière suivante :

Distribution génique	$\sum_i p'_i A_i$	$p'_i = \sum_j P'_{ij}$
Distribution génotypique	$\sum_{ij} P'_{ij} A_i A_j$	$P'_{ij} = P_{ij} \frac{m_{ij}}{\bar{m}}$

Remarque:

Lorsque l'on sélectionne sur la valeur génotypique, on peut considérer que m_{ij} est la valeur de G pour le génotype $A_i A_j$; m_{ij} est alors proportionnelle à g_{ij} .

Après sélection, à la nouvelle génération, on obtient une population dont les distributions génotypique et génique ont évolué:

$$\begin{aligned} \text{fréquence du génotype } A_i A_i : & \quad P'_{ii} = p_i^2 \frac{m_{ii}}{\bar{m}} \\ \text{fréquence du génotype } A_i A_j : & \quad P'_{ij} = 2p_i p_j \frac{m_{ij}}{\bar{m}} \end{aligned}$$

$$\begin{aligned} \text{fréquence de l'allèle } A_i : & \quad p'_i = P'_{ii} + \frac{1}{2} \sum_{j \neq i} P'_{ij} = \sum_j P'_{ij} = \sum_j P_{ij} \frac{m_{ij}}{\bar{m}} \\ & \quad = \frac{m_i \sum_j P_{ij}}{\bar{m}} = \frac{m_i p_i}{\bar{m}} \end{aligned}$$

L'évolution des fréquences alléliques d'une génération à la suivante s'écrit donc:

$$\Delta p_i = p'_i - p_i = p_i \left(\frac{m_i}{\bar{m}} - 1 \right) = p_i \frac{m_i - \bar{m}}{\bar{m}}$$

avec $m_i - \bar{m}$ l'excès moyen de l'allèle A_i

Ainsi, la fréquence d'un allèle augmente d'une génération à l'autre si sa valeur sélective est supérieure à la valeur sélective moyenne de la population, elle diminue dans le cas contraire.

L'évolution de la valeur sélective moyenne s'écrit :

$$\Delta \bar{m} = \bar{m}' - \bar{m} = \sum_{ij} P'_{ij} m_{ij} - \sum_{ij} P_{ij} m_{ij}$$

Devant cette formule, Fisher propose la démarche suivante:

Il choisit une approximation de m_{ij} par le modèle additif

$$m_{ij} = \bar{m} + \alpha_i + \alpha_j$$

cette quantité est obtenue en minimisant $\sum_{ij} P_{ij} (m_{ij} - \bar{m} - \alpha_i - \alpha_j)^2$

sous la contrainte $\sum_i P_i \alpha_i = 0$

La variance additive s'écrit alors : $\sigma_A^2 = 2 \sum_i p_i (m_i - \bar{m})^2 = 2\bar{m} \sum_i \Delta p_i \alpha_i$

Cette démarche est classique en analyse de variance et conduit à l'expression de $\Delta\bar{m}$ sous la forme :

$$\begin{aligned} \Delta\bar{m} &\simeq \sum_{i,j} (P'_{ij} - P_{ij})(\bar{m} + \alpha_i + \alpha_j) = \sum_{i,j} (P'_{ij} - P_{ij})(\alpha_i + \alpha_j) = 2 \sum_i \alpha_i \sum_j (P'_{ij} - P_{ij}) \\ &= 2 \sum_i \alpha_i \Delta p_i = \frac{\sigma_A^2}{\bar{m}} \end{aligned}$$

Ainsi les évolutions des fréquences alléliques et de la valeur sélective moyenne de la population s'écrivent :

$\Delta p_i = p'_i - p_i = p_i \frac{m_i - \bar{m}}{\bar{m}}$	$m_i = \frac{\sum_{ij} P_{ij} m_{ij}}{\sum_{ij} P_{ij}}$
$\Delta\bar{m} = \bar{m} - \bar{m}' \simeq \frac{\sigma_A^2}{\bar{m}}$	$\sigma_A^2 = \text{la variance additive de la valeur sélective}$

Pour mémoire :

Le théorème fondamental de la sélection a été proposé par Fisher sous la forme suivante:

$$\frac{\Delta\bar{m}}{\bar{m}} = \frac{\sigma_A^2}{\bar{m}^2}$$

“Le taux d'accroissement de la valeur sélective moyenne d'une population est égal, à tout instant, à la variance relative de la part additive des valeurs sélectives des divers génotypes.”

Cette variation relative de la valeur sélective moyenne caractérise la rapidité d'évolution de la population sous l'action des forces de sélection. Son intérêt est donc primordial. Malheureusement, l'évaluation de la part génétique additive d'un caractère aussi complexe que la valeur sélective ne peut être réalisée qu'au prix d'hypothèses peu réalistes.

Conclusion.

Pour comprendre les débats motivés par des interprétations variées des conditions dans lesquelles Fisher a travaillé, retenons que :

- Fisher ne mesure pas $\Delta\bar{m}$ comme l'évolution des moyennes des valeurs sélectives m_{ij} mais de leurs approximations $\bar{m} + \alpha_i + \alpha_j$. Il est clair que $\Delta\bar{m}$ n'est la mesure de l'évolution des valeurs sélectives moyennes que si le modèle additif est respecté.

- Fisher ne fait aucune hypothèse concernant le système de reproduction. Cependant, on montre que la formule de Fisher traduit bien $\Delta\bar{m}$ s'il y a panmixie. De plus, Kempthorne a généralisé ce résultat en montrant que cette formule demeurerait exacte en l'absence de panmixie si le produit des fréquences des homozygotes restait proportionnel au carré des fréquences des hétérozygotes. (Ce résultat a pour seul mérite d'élargir le domaine d'application de la formule de Fisher).
- Enfin, Fisher s'intéresse au $\Delta\bar{m}$ dû à la sélection au niveau d'une génération sans chercher à traduire une évolution continue de la population.

6.2 Généralisation au cas multiloci (Ewens, 1989)

On note: I un gamète. $I = (1_s, 2_u, \dots, k_v, \dots)$ k :locus et k_u :allèle u au k^{eme} locus
 P_{IJ} la fréquence du génotype (I, J)
 m_{IJ} sa valeur sélective
 p_{k_u} la fréquence de l'allèle u au locus k
 P_{k_u, l_v} la fréquence des gamètes portant simultanément l'allèle u au locus k et l'allèle v au locus l .

La variation de la fréquence P_{IJ} entre l'âge de conception et l'âge de reproduction de la génération parentale, c'est-à-dire après sélection, s'écrit:

$$\Delta P_{IJ} = P_{IJ} \frac{m_{IJ} - \bar{m}}{\bar{m}} \quad \text{avec} \quad \bar{m} = \sum_{I,J} P_{IJ} m_{IJ}$$

Cette formulation est identique à celle utilisée pour les fréquences alléliques. De la même façon :

$$\Delta p_{k_u} = \sum_{I,J} 1_{k_u} P_{IJ} \frac{m_{IJ} - \bar{m}}{\bar{m}}$$

$$\text{avec} \quad 1_{k_u} = \begin{cases} 1 & \text{si le génotype } (I, J) \text{ contient l'allèle } u \text{ au locus } k \\ 0 & \text{dans le cas contraire} \end{cases}$$

Reprenant l'approche de Fisher, Ewens propose de remplacer m_{IJ} par des approximations par le modèle additif définies en minimisant :

$$\sum_{I,J} P_{IJ} (m_{IJ} - \bar{m} - \sum_{I,J} 1_{k_u} \alpha_{k_u})^2 \quad \text{sous la contrainte} \quad \sum_u \alpha_{k_u} p_{k_u} = 0$$

et avec 1_{k_u} défini comme précédemment.

Cette opération est classique et l'on obtient, comme Fisher, l'expression de la variation de la représentation additive des valeurs sélectives sous la forme :

$$\Delta\bar{m} = \frac{\sigma_A^2}{\bar{m}}$$

Alors que dans le cas monolocus, la distinction entre \bar{m} et $\sum_{i,j} P_{ij}(\bar{m} + \alpha_i + \alpha_j)$ était purement conceptuelle, dans le cas multiloci, étant donné le *déséquilibre de liaison*, la covariance entre les effets des allèles à 2 loci différents d'un même gamète ne sera pas nulle et l'on aura:

$$\bar{m} \neq \sum_{i,j} P_{ij}(\bar{m} + \sum_{k,u}^{(i,j)} \alpha_{ku})$$

6.3 Passage au temps continu

Approche de Wright

Les résultats proposés par Fisher et Ewens concernent :

- . les valeurs sélectives approchées ($\bar{m} + \alpha_i + \alpha_j$)
- . l'évolution sur une génération isolée.

Wright tente de contourner ces deux contraintes en traduisant l'évolution des moyennes des valeurs sélectives en continu (chevauchement de générations).

Pour ce faire il se limite au cas biallélique. Dans une telle situation on a :

$$\begin{aligned}\bar{m} &= m_{11}p^2 + 2m_{12}pq + m_{22}q^2 = pm_1 + qm_2 \\ m_1 &= (m_{11}p^2 + m_{12}pq)/(p^2 + pq) = m_{11}p + m_{12}q \\ m_2 &= (m_{22}q^2 + m_{12}pq)/(q^2 + pq) = m_{22}q + m_{12}p\end{aligned}$$

avec m_{11} , m_{12} et m_{22} les valeurs sélectives des génotypes A_1A_1 , A_1A_2 et A_2A_2 ,
 p la fréquence de l'allèle A_1 et m_1 et m_2 les valeurs sélectives de A_1 et A_2 .

On s'intéresse à $\frac{dp}{dt}$ et $\frac{d\bar{m}}{dt}$ qui sont en situation continue les valeurs correspondant aux Δp et Δm calculées par Fisher et Ewens.

Faisons l'hypothèse que la population des allèles (A_1, A_2) soit de taille finie N répartie en N_1 éléments pour A_1 et N_2 pour A_2 .

$$N = N_1 + N_2 \quad N_1 = Np_1 \quad N_2 = Np_2$$

On peut admettre que les variations dN_i de N_i dans un intervalle de temps dt soient proportionnelles à l'effectif Np_i et à m_i .

ainsi
$$\frac{dN_i}{dt} = Np_i m_i$$

On a d'autre part :

$$\frac{dN}{dt} = \sum_i \frac{dN_i}{dt} = N \sum_i p_i m_i = N\bar{m}$$

et puisque $p_i = \frac{N_i}{N} \implies \frac{dN_i}{dt} = \frac{dN}{dt} p_i + N \frac{dp_i}{dt}$
 soit $\frac{dp_i}{dt} = \frac{1}{N} \left(\frac{dN_i}{dt} - \frac{dN}{dt} p_i \right)$ donc :

$$\boxed{\frac{dp_i}{dt} = p_i(m_i - \bar{m})}$$

Soit en revenant aux notations précédentes :

$$\boxed{\frac{dp}{dt} = p(1-p)[m_{11}p + m_{12}(1-2p) - m_{22}(1-p)]}$$

Parallèlement, il est facile de calculer $\frac{d\bar{m}}{dp}$:

$$\frac{d\bar{m}}{dp} = 2[m_{11}p + m_{12}(1-2p) - m_{22}(1-p)]$$

De plus :

$$\begin{aligned} \sigma_A^2 &= 2[p(m_1 - \bar{m})^2 + q(m_2 - \bar{m})^2] = 2p[q(m_1 - m_2)]^2 + 2q[p(m_2 - m_1)]^2 \\ &= 2pq(m_1 - m_2)^2 = 2pq[m_{11}p + m_{12}(1-2p) - m_{22}(1-p)]^2 \end{aligned}$$

En rassemblant ces résultats, on obtient :

$$\boxed{\frac{d\bar{m}}{dt} = \sigma_A^2}$$

On montre que ce résultat est généralisable au cas multiallélique.

Remarque :

Contrairement à Fisher, Wright fait l'hypothèse de la panmixie.

Généralisation de l'approche de Fisher

Ewens (1989) propose une généralisation du Théorème fondamental de la sélection au temps continu, en supposant que :

- un taux $P_{ij}(1 - m_{ij})dt$ des génotypes A_iA_j disparaîtront entre les instants t et $t + 1$,
- les individus morts seront remplacés par des nouveaux-nés parmi lesquels le génotype A_iA_j aura un taux d'accroissement P_{ij}^* ,

$$\frac{dP_{ij}}{dt} = P_{ij}(m_{ij} - 1) + (1 - \bar{m})P_{ij}^*$$

Si l'on suppose de plus que $p'_i = p_i^*$, c'est-à-dire que l'allèle A_i de la population naissante a la même fréquence que dans la population parente, alors :

$$\sum_{ij} \left(\frac{d}{dt} P_{ij} \right) (\bar{m} + \alpha_i + \alpha_j) = \sigma_A^2$$

Ainsi, dans le cas continu, le résultat de la formule de Fisher (en supposant le modèle additif) est identique à celui de la formule de Wright (en supposant la panmixie).

Enfin, Kimura donne, dans le cas continu, une généralisation du théorème de Fisher, sans supposer le modèle additif, avec la formule suivante :

$$\frac{d\bar{m}}{dt} = \sigma_A^2 + \sum_{ij} d_{ij} \frac{d}{dt} P_{ij}$$

avec $d_{ij} = m_{ij} - \bar{m} - \alpha_i - \alpha_j$

7 La Génétique Quantitative

7.1 Modèle infinitésimal et approche Gaussienne

En supposant la valeur adaptative d'un génotype constante dans le temps et une population de taille infinie, Lande (1976) décrit l'évolution dans le temps de la valeur moyenne d'un caractère par la formule suivante:

$$\Delta \bar{z}(t) = \bar{z}(t+1) - \bar{z}(t) = h^2(\bar{z}_w(t) - \bar{z}(t))$$

où $\bar{z}_w(t)$ est la valeur moyenne du caractère après sélection avant reproduction, $(\bar{z}_w(t) - \bar{z}(t)) = s$ reflète la *différentielle de sélection*, c'est-à-dire l'écart entre la valeur moyenne de la population après sélection et la valeur moyenne avant sélection. h^2 est l'héritabilité du caractère déterminée par le système génétique.

Lorsqu'on est en présence d'une caractérisation multicritère :

$$\bar{z}(t) = (\bar{z}_1(t), \bar{z}_2(t), \dots, \bar{z}_p(t))$$

En posant $P = G + E$ où G et E sont respectivement l'effet additif génétique et l'effet environnemental, avec G et E de distributions Normales et indépendantes, Lande (1979) réécrit sa formule en l'appliquant à l'ensemble des caractères i :

$$\Delta \bar{z} = GP^{-1}s$$

Soit la distribution des phénotypes, z , à la génération t avant sélection: $p(z, t)$

Le phénotype moyen avant sélection s'écrit:

$$\bar{z}(t) = \int_z z.p(z, t).dz$$

Soit la valeur adaptative (*Fitness*) d'un individu z , $m(z)$,
la *Fitness* moyenne dans la population est:

$$\bar{m} = \int_z p(z, t).m(z).dz$$

Le phénotype moyen après sélection s'écrit:

$$\bar{z}_w(t) = \frac{1}{\bar{m}} \int_z z.p(z, t).m(z).dz$$

On suppose alors une distribution Normale des caractères phénotypiques:

$$p(z, t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z - \bar{z}(t))^2}{2\sigma^2}} \quad \text{avec } \sigma^2 \text{ indépendante de } \bar{z}(t)$$

On en déduit la variation de *Fitness* moyenne relative à celle du phénotype moyen:

$$\begin{aligned}
\frac{\partial \bar{m}}{\partial \bar{z}(t)} &= \int_z \left(\frac{\partial p(z, t)}{\partial \bar{z}(t)} m(z) + p(z, t) \frac{\partial m(z)}{\partial \bar{z}(t)} \right) dt \\
&= \int_z \left(\frac{\partial p(z, t)}{\partial \bar{z}(t)} m(z) \right) dt \\
&= \int_z \frac{z(t) - \bar{z}(t)}{\sigma^2} p(z, t) m(z) dt \\
&= \frac{1}{\sigma^2} \left(\int_z z(t) p(z, t) m(z) dt - \int_z \bar{z}(t) p(z, t) m(z) dt \right) \\
&= \frac{\bar{m}}{\sigma^2} (\bar{z}_w(t) - \bar{z}(t))
\end{aligned}$$

On peut alors réécrire la première équation de la manière suivante :

$$\Delta \bar{z}(t) = \frac{h^2 \sigma^2}{\bar{m}} \frac{\partial \bar{m}}{\partial \bar{z}(t)} = h^2 \sigma^2 \frac{\partial \ln \bar{m}}{\partial \bar{z}(t)} = \sigma_A^2 \frac{\partial \ln \bar{m}}{\partial \bar{z}(t)}$$

Si l'on note l'opérateur gradient $\nabla = \left(\frac{\partial}{\partial \bar{z}_1}, \dots, \frac{\partial}{\partial \bar{z}_k} \right)'$

$\nabla \ln \bar{m}$ est le *gradient de sélection* et on peut écrire :

$$\Delta \bar{z} = G \sum_i^k \frac{\partial \ln \bar{m}}{\partial \bar{z}_i} = G \nabla \ln \bar{m} = G P^{-1} s$$

où G est la matrice de variance-covariance génétique additive, P la matrice de variance-covariance phénotypique et s la différentielle de sélection.

7.2 Lien avec le théorème de Fisher (Charnov 1989)

Le modèle quantitatif est en fait une approximation linéaire du théorème fondamental de la sélection (Fisher).

Cas unicritère

On considère un développement de $m(z)$ et son approximation du premier ordre :

$$m(z) \simeq m_L(z) = m(\bar{z}) + (z - \bar{z}) \frac{\partial m}{\partial z}(\bar{z})$$

Remarque :

$$* \quad m_L(\bar{z}) = m(\bar{z}) \quad \text{donc} \quad \frac{\partial m_L}{\partial z}(\bar{z}) = \frac{\partial m}{\partial z}(\bar{z})$$

- * m_L est une variable aléatoire qui a pour moyenne $m(\bar{z})$ et pour variance $VAR(m_L(z)) = var(z) \cdot \left(\frac{\partial m}{\partial z}(\bar{z}) \right)^2$
- * Si l'on écrit $z = z_g + z_e$ où z_g est la partie génétique additive et si l'on fait les hypothèses habituelles : $var(z) = var(z_g) + var(z_e)$ et $var(z_g) \left(\frac{\partial m}{\partial z}(\bar{z}) \right)^2$ représente la partie de variance génétique de $m_L(z)$.

Si l'on suppose que \bar{z}_{t+1} est peu différent de \bar{z}_t en réécrivant l'approximation du premier ordre du développement de $m(z)$, on a :

$$m_L(\bar{z}_{t+1}) = m_L(\bar{z}_t) + (\bar{z}_{t+1} - \bar{z}_t) \underbrace{\frac{\partial m}{\partial z}(\bar{z}_t)}_{\frac{\partial m_L}{\partial z}(\bar{z}_t)}$$

d'où $\Delta \bar{m}_L = \Delta \bar{z} \frac{\partial m_L}{\partial z}(\bar{z}_t)$

Or d'après Fisher on a :

$$\Delta \bar{m}_L = \frac{V_g(m_L(z))}{\bar{m}_L} \simeq \Delta \bar{m} = \Delta(\bar{z}) \frac{\partial m_L}{\partial z}(\bar{z}_t) = \frac{var(z_g)}{\bar{m}_L} \left(\frac{\partial m}{\partial z}(\bar{z}_t) \right)^2$$

D'où $\Delta z = \frac{var(z_g)}{\bar{m}_L} \frac{\partial m}{\partial z}(\bar{z})$

On retrouve bien ici la formule de Lande.

Cas multicritère

Soit la valeur sélective $M(Z_1, \dots, Z_k)$, une variable aléatoire fonction de plusieurs variables aléatoires. Les Z_i sont les variables phénotypiques associées à différents caractères. On réalise un développement de Taylor au premier ordre de la valeur sélective d'un individu autour de la valeur moyenne de Z (\bar{Z}_i):

$$m_L(z) = m(\bar{z}_1, \dots, \bar{z}_k) + \sum_{i=1}^k \frac{\partial m}{\partial z_i}(z_i - \bar{z}_i)$$

Sous forme vectorielle, l'expression devient:

$$m_L(z) = m(\bar{z}) + (z - \bar{z})' \nabla m(z)$$

La moyenne de cette approximation linéaire $\bar{m}_L = m(\bar{z})$.

La variance de m_L peut être calculée:

$$\begin{aligned} V(\dot{m}_L) &= E[\nabla m'(z - \bar{z})(z - \bar{z})' \nabla m] \\ &= \nabla m' E[(z - \bar{z})(z - \bar{z})'] \nabla m \\ &= \sum_i^k V(z_i) \left(\frac{\partial m}{\partial z_i} \right)^2 + \sum_{i \neq j}^k cov(z_i, z_j) \left(\frac{\partial m}{\partial z_i} \right) \left(\frac{\partial m}{\partial z_j} \right) \end{aligned}$$

$$= \sum_{i=1}^k \left(\frac{\partial m}{\partial z_i} \right) \left(V(z_i) \frac{\partial m}{\partial z_i} + \sum_{i \neq j}^k \text{cov}(z_i, z_j) \frac{\partial m}{\partial z_j} \right)$$

Si l'évolution de la valeur sélective moyenne d'une population \bar{m}_L due à la sélection est lente, on pourra réaliser un développement limité de Taylor au premier ordre de la valeur sélective moyenne après sélection autour de la valeur sélective moyenne avant sélection:

$$\bar{L}_L(t+1) \simeq \bar{m}_L(t) + \sum_i^k \frac{\partial \bar{m}_L}{\partial \bar{z}_i} \Delta \bar{z}_i$$

La fonction $m(\bar{z}_1, \dots, \bar{z}_k)$ aura donc exactement la même forme que la fonction $m(z_1, \dots, z_k)$. La variance de la valeur sélective moyenne s'écrira donc:

$$V(\bar{m}_L) = \sum_{i=1}^k \left(\frac{\partial \bar{m}}{\partial \bar{z}_i} \right) \left(V(z_i) \frac{\partial \bar{m}}{\partial \bar{z}_i} + \sum_{i \neq j}^k \text{cov}(z_i, z_j) \frac{\partial \bar{m}}{\partial \bar{z}_j} \right)$$

Par ailleurs, si l'on suppose que chaque Z_i est la somme de deux variables aléatoires indépendantes:

$$Z_i = Z_{ig} + Z_{ie}$$

avec Z_{ig} la composante génétique additive,

Z_{ie} la composante restante (environnementale, dominance, etc.),

$E(Z_{ie}) = \bar{Z}_{ie} = 0$ et $\text{cov}(Z_{ig}, Z_{ie}) = 0$

alors les variances et covariances intervenant dans l'équation de $V(\bar{m}_L)$ sont de type génétique additif (σ_A^2).

Or Fisher écrit l'évolution de la valeur sélective moyenne de la manière suivante:

$$\bar{m}(t+1) - \bar{m}(t) = \frac{V(\bar{m}(t))}{\bar{m}(t)}$$

On obtient alors:

$$\bar{m}_L(t+1) - \bar{m}_L(t) = \frac{V(\bar{m}_L(t))}{\bar{m}_L(t)}$$

Il suffit enfin de résoudre: $\frac{V(\bar{m}_L)}{\bar{m}_L} = \sum_{i=1}^k \frac{\partial \bar{m}_L}{\partial \bar{z}_i} \Delta \bar{z}_i$

$$\iff \bar{m} \Delta \bar{z}_i = V(z_i) \frac{\partial \bar{m}}{\partial \bar{z}_i} + \sum_{i \neq j}^k \text{cov}(z_i, z_j) \frac{\partial \bar{m}}{\partial \bar{z}_j}$$

ou encore sous forme matricielle:

$$\begin{pmatrix} \bar{m} \Delta \bar{z}_1 \\ \vdots \\ \bar{m} \Delta \bar{z}_k \end{pmatrix} = \begin{pmatrix} V(z_1) & cov(z_1 z_2) & \dots \\ cov(z_1 z_2) & V(z_2) & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \partial \bar{m} / \partial \bar{z}_1 \\ \vdots \\ \partial \bar{m} / \partial \bar{z}_k \end{pmatrix}$$

$$\Leftrightarrow \boxed{\Delta \bar{z} = G \frac{1}{\bar{m}} \frac{\partial \bar{m}}{\partial \bar{z}} = G \frac{\partial \ln(\bar{m})}{\partial \bar{z}} = G \nabla \ln \bar{m}}$$

La formule de Lande s'obtient donc en linéarisant le théorème de Fisher. La prédiction linéaire ne nécessite donc pas forcément d'avoir recours à l'hypothèse gaussienne mais uniquement à toute raison justifiant une écriture au premier ordre.

7.3 Généralisation aux cas d'appariements préférentiels

La prise en compte d'un système de reproduction non panmictique dans l'évolution de la valeur phénotypique moyenne d'une population sous sélection est proposée par Perrin (1993).

Soit une fonction symétrique $f(z_1, z_2) = z'$ représentant la valeur moyenne d'un enfant du couple (z_1, z_2) ,

S'il y a panmixie la distribution (z_1, z_2) s'écrit :

$$\varphi(z_1, z_2) = \frac{m(z_1)}{\bar{m}} \frac{m(z_2)}{\bar{m}} p(z_1) p(z_2)$$

Pour traduire que l'appariement est préférentiel on introduit une fonction de densité $a(z_1, z_2)$ qui dépend du mode d'appariement et vaut 1 dans le cas de la panmixie.

La distribution s'écrit alors : $\varphi(z_1, z_2) = a(z_1, z_2) \frac{m(z_1)}{\bar{m}} p(z_1) \cdot \frac{m(z_2)}{\bar{m}} p(z_2)$

La valeur phénotypique moyenne à la nouvelle génération peut alors être calculée de la manière suivante:

$$\bar{z}' = \int_{z_1} \int_{z_2} z' \varphi(z_1, z_2) dz_1 dz_2 = \frac{1}{\bar{m}^2} \int_{z_1} \int_{z_2} f(z_1, z_2) \cdot a(z_1, z_2) \cdot m(z_1) \cdot m(z_2) \cdot p(z_1) \cdot p(z_2) \cdot dz_1 \cdot dz_2$$

En utilisant le développement linéaire de m ($m_L(z) = m(\bar{z}) + (z - \bar{z})' \nabla m(\bar{z})$) et sachant que $m_L(\bar{z}) = \bar{m}$ et donc que $\nabla m_L(\bar{z}) = \nabla \bar{m}$ on obtient:

$$\begin{aligned} \bar{z}' &= \frac{1}{\bar{m}^2} \int_{z_1} \int_{z_2} f(z_1, z_2) \cdot a(z_1, z_2) \cdot (m(\bar{z}) + (z_1 - \bar{z})' \nabla \bar{m}) \\ &\quad \cdot (m(\bar{z}) + (z_2 - \bar{z})' \nabla \bar{m}) \cdot p(z_1) \cdot p(z_2) \cdot dz_1 \cdot dz_2 \\ &= \int_{z_1} \int_{z_2} f(z_1, z_2) \cdot a(z_1, z_2) \cdot p(z_1) \cdot p(z_2) \cdot dz_1 \cdot dz_2 \end{aligned}$$

$$\begin{aligned}
& + \frac{2}{\bar{m}} \nabla \bar{m} \int_{z_1} \int_{z_2} f(z_1, z_2) \cdot a(z_1, z_2) \cdot (z_1 - \bar{z}) \cdot p(z_1) \cdot p(z_2) \cdot dz_1 \cdot dz_2 \\
& + \left(\frac{1}{\bar{m}} \nabla \bar{m} \right)^2 \int_{z_1} \int_{z_2} f(z_1, z_2) \cdot a(z_1, z_2) \cdot (z_1 - \bar{z}) \cdot (z_2 - \bar{z}) \cdot p(z_1) \cdot p(z_2) \cdot dz_1 \cdot dz_2
\end{aligned}$$

Rendons plus explicites les trois termes de cette équation:

- Le 1^{er} terme est la valeur moyenne \bar{z}_0 en descendance contrôlée de la population de départ (sans sélection); il traduit l'effet du système de reproduction sur la moyenne. On le note $\bar{z} + \Delta_m \bar{z}$, (chez les diploïdes, en l'absence d'épistasie, $\Delta_m \bar{z} = \bar{z}_0 - \bar{z} = 0$).
- Le 2^{me} terme est la covariance parent-enfant en descendance contrôlée sur les deux sexes (père et mère) $cov_m(PO)$.
- Le 3^{me} terme est la covariance entre la valeur des enfants (O) et celle des deux parents (P_1 et P_2), relativement au mode de reproduction choisi. On note cette trivariance: $cov_m(P_1, P_2, O)$.

L'évolution de la valeur phénotypique moyenne d'une population sous l'action combinée de la **sélection** et de la **reproduction** $\bar{z}' - \bar{z}$ s'écrit:

$$\underbrace{\bar{z}' - \bar{z} = \Delta \bar{z} = \Delta_m \bar{z} + 2cov_m(PO) \nabla \ln \bar{m} + cov_m(P_1, P_2, O) (\nabla \ln \bar{m})^2}_{\Delta \bar{z}}$$

En utilisant la formule de Lande on retrouve ici les premiers termes de la formule classique du progrès génétique réalisé au niveau d'une population ΔG_P (Gallais, 1990):

$$\boxed{\Delta G_P = \mu_{n+1} - \mu_n = \underbrace{\mu_{n,1} - \mu_n}_{\Delta G_M} + i\theta \underbrace{\frac{cov(TM)}{\sqrt{varT}}}_{\Delta G_S}}$$

où $\mu_{n,1}$ (notée \bar{z}_0) est la moyenne de la population $P_{n,1}$ issue par intercroisement sans sélection de la population P_n .

$cov(TM)$ est la covariance parent-enfant et $varT$, la variance phénotypique des parents.

$\Delta G_M = (\mu_{n,1} - \mu_n)$ est l'effet du *système de reproduction* sans sélection.

$\Delta G_S = i\theta \rho_{TM} h_T \sigma_M$ est l'effet de la *sélection*,

h_T^2 étant l'hérabilité du système de test T ,

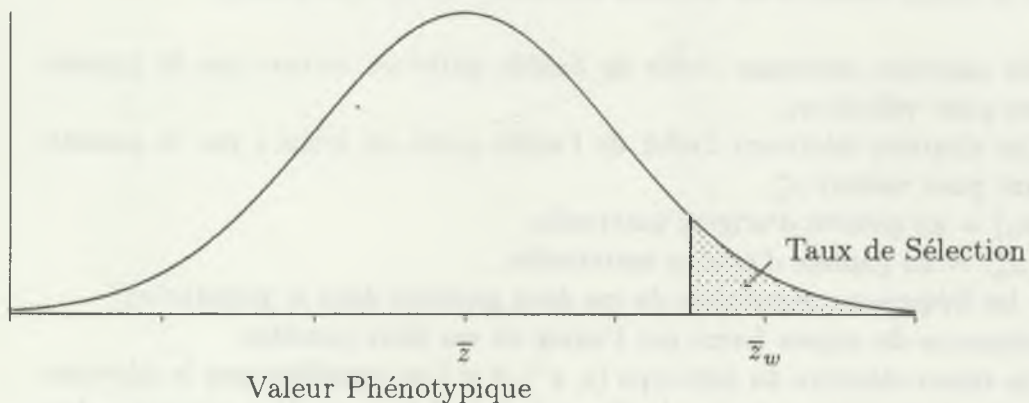
ρ_{TM} la corrélation génétique entre la valeur des parents selon le système de test T et la valeur des enfants M ,

σ_M^2 la variance génétique des descendance des unités sélectionnées M .

θ est le degré de contrôle de la sélection sur les deux sexes ($\theta = 1$ ou 2).

i est l'intensité de sélection définie par la différentielle de sélection en unités d'écart-type de la distribution des valeurs phénotypiques (voir figure ci-dessous).

Distribution de densité



i dépend du pourcentage d'individus sélectionnés τ (*taux de sélection*); il existe des tables donnant la correspondance entre ces deux paramètres, établies à partir des propriétés de la loi normale.

Sachant que la *différentielle de sélection* $s = \bar{z}_w - \bar{z}$:

$$i = \frac{s}{\sqrt{\text{var}T}}$$

La question est donc de savoir quelle est la nature de l'approximation faite en négligeant le troisième terme de l'équation générale. Seule la double hypothèse de panmixie et de régression linéaire enfant-[parent1 parent2] ($O = \alpha P_1 + \beta P_2$) permet d'annuler ce terme complémentaire.

8 Formalisation de Turelli et Barton (1990)

Indépendamment de la prise en compte d'appariements préférentiels, l'évolution de la moyenne dépend des moments d'ordres plus élevés. Si l'on refuse l'hypothèse Gaussienne il est important de chercher à caractériser la distribution de ces moments. Barton et Turelli ont écrit la valeur sélective en fonction des moments de tous ordres.

On pose:

X_i = la variable aléatoire décrivant l'effet de l'allèle porté au locus i par le gamète paternel, prenant pour valeurs x_i ,

X_i^* = la variable aléatoire décrivant l'effet de l'allèle porté au locus i par le gamète maternel, prenant pour valeurs x_i^* ,

$x = (x_1, x_2, \dots, x_n)$ = un gamète d'origine paternelle,

$x^* = (x_1^*, x_2^*, \dots, x_n^*)$ = un gamète d'origine maternelle,

$p(x)$ et $p(x^*)$ = les fréquences respectives de ces deux gamètes dans la population,

$P(x, x^*)$ = la fréquence du zygote formé par l'union de ces deux gamètes,

Si $m(x, x^*)$ est la valeur sélective du génotype (x, x^*) et si l'on considère que la sélection ne porte que sur la viabilité, la variation $\Delta_S P(x, x^*)$ de la fréquence de ce génotype due à la sélection s'écrit:

$$\Delta_S P(x, x^*) = P(x, x^*) \frac{m(x, x^*) - \bar{m}}{\bar{m}} \quad (1)$$

avec $\bar{m} = \sum_{x, x^*} m(x, x^*) \cdot P(x, x^*)$

On note au niveau de la population:

Z = la valeur phénotypique et G = la valeur génotypique,

ε = la variable aléatoire environnementale, indépendante de G , de moyenne nulle et de variance V_e ,

$V = \text{var}(Z)$ la variance phénotypique et $V_g = \text{var}(G)$ la variance génétique additive,

$M_{k,p} = E((Z - \bar{Z})^k)$ pour $k \geq 3$: moment phénotypique,

$M_{k,g} = E((G - \bar{G})^k)$ pour $k \geq 3$: moment génotypique,

\bar{Z}_t = moyenne phénotypique avant sélection à la génération t ,

\bar{Z}'_t = moyenne phénotypique après sélection à la génération t ,

$\Delta_S \bar{Z} = \bar{Z}'_t - \bar{Z}_t$ = variation due à la sélection,

$\Delta \bar{Z} = \bar{Z}_{t+1} - \bar{Z}_t$ = variation d'une génération à la suivante.

Les hypothèses d'additivité du modèle permettent d'écrire la valeur phénotypique:

$$Z = \sum_i (X_i + X_i^*) + \varepsilon = G + \varepsilon$$

La moyenne phénotypique dans la population s'écrit:

$$\bar{Z} = E(Z) = E(G) + 0 = \sum_i (E(X_i) + E(X_i^*)) = \sum_i (m_i + m_i^*)$$

En supposant les caractères non liés au sexe: $m_i = m_i^*$

et les accouplements au hasard: $\Delta m_i = \Delta_S m_i$ (la variation d'une génération à l'autre des effets moyens des allèles portés par le locus i est uniquement due à la sélection et non au système de reproduction), on peut écrire:

$$\begin{aligned}\Delta \bar{Z} &= \bar{Z}_{t+1} - \bar{Z}_t = \sum_i (\Delta m_i + \Delta m_i^*) \\ &= \sum_i (\Delta_S m_i + \Delta_S m_i^*) = 2 \sum_i \Delta_S m_i\end{aligned}$$

En sommant sur les génotypes, $m_i = \sum_{x, x^*} x_i \cdot P(x, x^*)$.

La variation due à la sélection de la moyenne dans la population des effets des allèles portés au locus i s'écrit:

$$\Delta_S m_i = \sum_{x, x^*} x_i \cdot \Delta_S P(x, x^*) \quad (2)$$

On considère alors un ensemble complet de fonctions des fréquences de tous les génotypes possibles, à savoir:

- . l'ensemble des moyennes m_i et m_i^* des effets des allèles à tous les locus i ,
- . l'ensemble des moments centrés de tous ordres des effets des allèles à tous les locus. il s'agit:

→ des variances:

$$C_{ii} = E[(X_i - m_i)^2] \text{ pour tous les locus } i,$$

→ des covariances:

$$C_{ij} = E[(X_i - m_i)(X_j - m_j)] \text{ (même gamète),}$$

$$C_{ij} = E[(X_i - m_i)(X_j^* - m_j^*)] \text{ (gamètes différents),}$$

pour tous les couples de loci (i, j)

→ et plus généralement, les moments centrés de tous ordres:

$$C_{ST} = E \left[\prod_{i \in S} (X_i - m_i) \cdot \prod_{j \in T} (X_j^* - m_j^*) \right]$$

pour tous sous-ensembles S et T (éventuellement vides) de l'ensemble des loci, les allèles aux locus de S provenant du gamète mâle et ceux de T , du gamète femelle.

La première équation (1) peut alors s'écrire:

$$\Delta_S P(x, x^*) = \sum_{y, y^*} Q[(x, x^*), (y, y^*)] \frac{\delta \ln \bar{m}}{\delta P(y, y^*)}$$

avec:

$$Q[(x, x^*), (y, y^*)] = \begin{cases} P(x, x^*) \cdot [1 - P(y, y^*)] & \text{si } (x, x^*) = (y, y^*) \\ -P(x, x^*) \cdot P(y, y^*) & \text{sinon} \end{cases}$$

En remplaçant dans l'équation (2), on obtient:

$$\Delta_S m_i = \sum_{x, x^*} x_i \sum_{y, y^*} Q[(x, x^*), (y, y^*)] \frac{\delta \ln \bar{m}}{\delta P(y, y^*)} \quad (3)$$

En utilisant l'ensemble complet de fonctions défini précédemment, la différentielle partielle est calculable si l'on connaît tous les $P(y, y^*)$. On note H_α , l'une quelconque des fonctions de cet ensemble de moments, alors:

$$\frac{\delta \ln \bar{m}}{\delta P(y, y^*)} = \sum_{\alpha} \frac{\delta \ln \bar{m}}{\delta H_{\alpha}} \frac{\delta H_{\alpha}}{\delta P(y, y^*)}$$

$\frac{\delta \ln \bar{m}}{\delta H_{\alpha}}$ est le gradient de sélection mesurable si H_{α} est mesurable.

$\frac{\delta H_{\alpha}}{\delta P(y, y^*)}$ est calculable d'après la définition de H_{α} .

En remplaçant dans l'équation (3), on obtient:

$$\begin{aligned} \Delta_S m_i &= \sum_{x, x^*} x_i \sum_{y, y^*} Q[(x, x^*), (y, y^*)] \sum_{\alpha} \frac{\delta \ln \bar{m}}{\delta H_{\alpha}} \frac{\delta H_{\alpha}}{\delta P(y, y^*)} \\ &= \sum_{\alpha} \left\{ \sum_{x, x^*} \sum_{y, y^*} x_i \cdot Q[(x, x^*), (y, y^*)] \frac{\delta H_{\alpha}}{\delta P(y, y^*)} \right\} \frac{\delta \ln \bar{m}}{\delta H_{\alpha}} \end{aligned}$$

La variation de m_i due à la sélection s'écrit donc comme une combinaison linéaire des gradients de sélection (dérivées partielles de $\ln \bar{m}$ pour toutes les fonctions H_{α}). Les coefficients (termes entre crochets) s'expriment en fonction des différents moments $C_{S,T}$.

Soit les gradients de sélection:

$$L_1 = \frac{\delta \ln \bar{m}}{\delta G} \quad L_2 = \frac{\delta \ln \bar{m}}{\delta V_g} \quad L_3 = \frac{\delta \ln \bar{m}}{\delta M_{k,g}} \text{ pour } k \geq 3$$

Si on se limite aux moments d'ordre inférieur ou égal à trois:

$$\begin{aligned} \Delta \bar{Z} &= 2 \sum_i \Delta_S m_i \\ &= 2 \left[\sum_{ij} C_{ij} L_1 + \sum_{ijk} C_{ijk} L_2 + \sum_{ijkh} K_{ijkh} L_3 + \dots \right] \end{aligned}$$

avec $K_{ijkh} = C_{ijkh} - C_{ij}C_{kh} - C_{ik}C_{jh} - C_{ih}C_{jk}$

Si l'on s'intéresse non plus à l'évolution de la moyenne phénotypique mais de la covariance entre les effets de deux loci d'un même gamète, on écrira:

$$\begin{aligned} \Delta_S C_{ij} &= \sum_{x, x^*} (x_i - m_i)(x_j - m_j) \Delta_S f(x, x^*) \\ &= \sum_k C_{ijk} L_1 + \sum_{k,h} (C_{ijkh} - C_{ij}C_{kh}) L_2 + \dots - \Delta_S m_i \Delta_S m_j \end{aligned}$$

Sous l'hypothèse d'appariements au hasard, les m_i ne sont pas modifiés par l'étape de reproduction: $\Delta m_i = m_i(t+1) - m_i(t) = \Delta_S m_i$

Par contre, la recombinaison doit être prise en compte pour le passage à la génération suivante en ce qui concerne les moments d'ordre supérieur. Ainsi:

$$C_{ij}(t+1) = (1 - r_{ij})C'_{ij}(t) + r_{ij}C_{i,j}(t)$$

avec r_{ij} le taux de recombinaison entre les loci i et j . On écrira la variation de C_{ij} d'une génération à la suivante:

$$\begin{aligned}\Delta C_{ij} &= C_{ij}(t+1) - C_{ij}(t) \\ &= C_{ij}(t+1) - C'_{ij}(t) + C'_{ij}(t) - C_{ij}(t) \\ &= C_{ij}(t+1) - C'_{ij}(t) + \Delta_S C_{ij}\end{aligned}$$

On obtient finalement:

$$\begin{aligned}\Delta C_{ij} &= \sum_k C_{ijk} L_1 + \sum_{k,h} (C_{ijkh} - C_{ij} C_{kh}) L_2 \\ &\quad - \Delta_S m_i \Delta_S m_j - r_{ij} C_{ij} \\ &\quad - r_{ij} \left(\sum_k C_{ijk} L_1 + \sum_{k,h} K_{ijkh} L_2 \right)\end{aligned}$$

L'ensemble de ces formules donnent les bases pour construire les équations de récurrence pour un jeu de paramètres décrivant complètement le système. Par ailleurs, l'équation associée à chaque moment fait intervenir des moments d'ordre supérieur. Cette écriture débouche donc sur un système d'équations explosif dont on ne peut tirer une solution dans le cas général. Pour obtenir les résultats classiques de la génétique quantitative, on a recours aux hypothèses de normalité des distributions valeurs phénotypiques et génotypiques.

Au niveau phénotypique, cette hypothèse permet d'estimer les coefficients de régression Parent-Enfant. On retrouve alors l'équation suivante:

$$R = \beta S$$

avec R : la réponse à la sélection et S : la différentielle de sélection, qui permet de prédire la moyenne des descendants sur une seule génération. Pour interpréter le β comme l'héritabilité, il faut alors faire des hypothèses génétiques, d'où la deuxième hypothèse de normalité de distribution des valeurs génotypiques. Or c'est à ce niveau que l'approximation devient très contestable:

- d'un point de vue mathématique, si la distribution des valeurs génotypiques est Normale sans sélection, elle ne l'est évidemment plus avec sélection,
- d'un point de vue génétique, le modèle infinitésimal considère un déterminisme génétique des caractères irréaliste : les locus sont nécessairement liés entre eux.

ANNEXES

Heritability: One Word, Three Concepts

A. Jacquard

Institut National d'Etudes Démographiques, 27 rue du Commandeur,
75675 Paris Cedex 14, France

SUMMARY

The term 'heritability', which evokes the image of transmission from parents to children, is used in biology to characterize the resemblance of related individuals in terms of a given characteristic, and to analyse the genetic and environmental causes of this resemblance. In fact, there are three definitions of heritability and it is important to distinguish among these carefully in order to avoid misinterpretations. Various techniques for measuring associated parameters are linked to these definitions. A rigorous analysis of the assumptions which permit the interpretation of parameter estimates is necessary to avoid false conclusions.

1. Definitions

1.1 *Heritability in Biometry: a Tentative Measure of 'Resemblance'*

We consider a quantitative character C measured by the variate X , and study its distribution in a defined population. Let p_i be the proportion of individuals for whom X has the value x_i . The mean μ and the variance of X in the population are given by

$$\mu = \sum_i p_i x_i, \quad \text{var}(X) = \sum_i p_i (x_i - \mu)^2.$$

Our purpose is to characterize the influence of one parent (say, the father) on the offspring (say, the son) for the character C .

Consider the set of fathers for whom X takes the value x_i ; for their sons the variate X follows a distribution with a conditional mean \bar{X}_i and conditional variance V_i . The link between fathers and sons may be described in two ways:

- (i) by considering the \bar{X}_i . If \bar{X}_i lies between x_i and μ , there is a *resemblance* between father and son: above-average fathers have, in the mean, above-average sons;
- (ii) by considering the V_i . If, on the average, V_i is less than $\text{var}(X)$, information about X for the father reduces the uncertainty about X for the son. The smaller the mean \bar{V}_i of the conditional variances, the more closely the value of C for the son is determined by its value for the father. The standard way of describing this determination is by defining a 'coefficient of *determination*',

$$D = 1 - \frac{\bar{V}_i}{\text{var}(X)}. \quad (1)$$

Note that *resemblance* as defined above implies *determination*, but the reverse is not true. Let us assume, for example, that, when X takes the value x_i for a father, the value for his son is $\mu + k(x_i - \mu)$, and the variances V_i are nil. Then the *determination* is total, yet there is *resemblance* only in the case $k > 0$ (and it may well be that $k < 0$, corresponding to systematic dissimilarity).

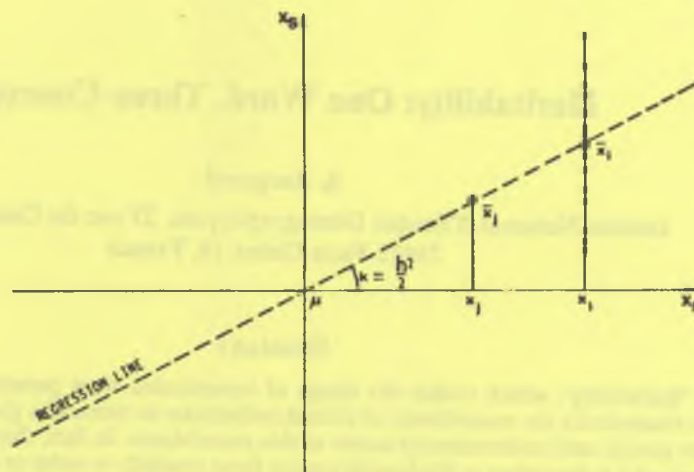


Figure 1. Regression of the value of a character, X_S , measured for a son on the value, X_F , measured for the father.

The standard relation between 'resemblance' and 'determination' is easy to demonstrate. Let X and Y be two random variables defined for the same experiment; to each value of Y there corresponds a conditional distribution of X , with a conditional mean and a conditional variance. The total variance of X is the sum of two terms, the variance of the conditional means and the mean of the conditional variances.

This result is the direct consequence of the definitions of the conditional means and variances and requires no other hypothesis. Applied to our problem, this relation gives

$$\text{var}(X) = \text{var}(\bar{X}_i) + \bar{V}_i$$

and

$$D = \frac{\text{var}(\bar{X}_i)}{\text{var}(X)}. \quad (2)$$

Let us now make three assumptions which are essential for the subsequent development:

- (i) The conditional means \bar{X}_i are linear functions of x_i , i.e.

$$\bar{X}_i = \mu + k(x_i - \mu); \quad (3)$$

see Fig. 1.

- (ii) The parameter k , the slope of the 'regression line' on which the (x_i, \bar{X}_i) lie, is positive (i.e. we exclude the case of systematic dissimilarity). Thus, k is a measure of resemblance: if $k = 0$ $\bar{X}_i = \mu$, the character of the son is not affected on the average by that of the father, and the 'resemblance' is nil; if $k = 1$, $\bar{X}_i = x_i$, and the 'resemblance' is perfect.

To simplify the notation, we assume also that

- (iii) μ and $\text{var}(X)$ are the same in the two successive generations which we study.

Under these assumptions, we can characterize the resemblance between son and father and the determination of sons by their fathers by the same parameter, since

$$\text{var}(\bar{X}_i) = \text{var}\{k(x_i - \mu)\} = k^2 \text{var}(X)$$

and, from (2),

$$D = k^2; \quad (4)$$

the slope of the regression line is thus equal to the square root of the coefficient of determination.

These two concepts may then be confused and named by the same word; classically this word is 'heritability' and the parameter that measures the heritability is written h^2 . Taking into account the fact that each offspring has two parents, h^2 is defined as equal to twice the parameter k which measures the resemblance between one offspring and one parent; thus,

$$h^2 = 2k = 2D^1 = 2 \left\{ \frac{\text{var}(\bar{X}_i)}{\text{var}(X)} \right\}^{\frac{1}{2}}. \quad (5)$$

If, instead of considering the value of the character in one parent, we had considered its mean value in the set of the two parents, the variance of the 'parental mean' would have been equal to $\frac{1}{2}V$ (assuming that the parents are uncorrelated), and the slope of the regression line would have been $2k$, or h^2 . Thus, we have the following

Definition. The 'heritability' of a character in a population is the slope of the linear regression line (if it exists) of the measurements of the character amongst children on the mean of the measurements of the character for their two parents.

The only underlying hypothesis is the linearity of the conditional means, i.e. the existence of a linear regression. In practice, the regression seldom deviates very far from linearity and the hypothesis is generally sufficiently realistic.

The heritability thus defined characterizes the relation between the values ascertained amongst offspring and those ascertained amongst parents. It is a measure based on observation which allows the prediction of the offspring values from those of the parents; the accuracy of the prediction increases with h^2 ; but this empirical heritability provides no explanation whatsoever for the observed resemblance. Research into causes is not the goal here. So far we have not made any reference to genetics. The situation is different for the types of heritability defined in §1.2 and §1.3; for these two 'genetic heritabilities' the determination of causes is indeed the goal.

Remark 1. In most real cases, the observed regressions are almost linear; this fact may be related to the Central Limit Theorem (especially its formulations by Lindeberg or Liapounov: if two characteristics are formed by the addition of many independent causes, their distribution is asymptotically bivariate normal, by the Central Limit Theorem; the conditional expectation of one variable is then a linear function of the other. This situation holds for most variables measured on living beings.

Remark 2. Knowledge of h^2 permits two calculations: one concerning the conditional mean, the other concerning the conditional variance. For example, if $h^2 = 0.80$ and if the character C has mean $\mu = 100$ and variance $V = 225$ [i.e. standard deviation $\sigma = 15$, as in the case of the intelligence quotient (IQ)], we can conclude that children of a parent with a value of 120 have a phenotype distribution with a conditional mean $= 100 + \frac{1}{2}(0.80)(120-100) = 108$, and a conditional variance $= 225 [1 - \{\frac{1}{2}(0.80)\}^2] = 189$; i.e. a conditional standard deviation $\sigma_c = 13.6$, only 9% less than the value in the population despite the seemingly high heritability.

The way in which h^2 has been defined may present a false picture of the importance of the resemblance between parents and offspring. The information on the offspring given by the information concerning one parent may be characterized by the ratio σ_c/σ of the conditional and total standard deviation:

$$\frac{\sigma_c}{\sigma} = \{1 - (\frac{1}{2}h^2)^2\}^{\frac{1}{2}} \approx 1 - \frac{1}{4}(h^2)^2.$$

Thus a heritability of 0.50 reduces the standard deviation amongst offspring by only 3%.

1.2 Genetic Heritability in the Broad Sense (or Coefficient of Genetic Determination): a Tentative Analysis of the 'Causes' of Resemblance

Any character manifested by an individual is the result of two categories of 'causes': the genetic information transmitted to the individual at the time of conception, and the contribution of the 'environment', in all its forms, during development.

The dispersion of the phenotypic values X of the character C amongst individuals belonging to a given population arises from individual differences in their genetic heritage and in their environmental conditions. It is tempting to try to split the variance of X into a part due to environmental variability and a part due to genetic variability:

$$\text{var}(X) = \text{var}_G(X) + \text{var}_E(X). \quad (6)$$

The relative 'part' of genetic causes in the total variability is named 'genetic heritability in the broad sense' and the parameter which measures it is written h_L^2 :

$$h_L^2 = \frac{\text{var}_G(X)}{\text{var}(X)}.$$

Unfortunately, such a relation has meaning only under very restrictive assumptions. This can be seen as follows: suppose that, in the population studied, there exist r genotypes, $g_1, \dots, g_i, \dots, g_r$, and s environments, $e_1, \dots, e_k, \dots, e_s$. Designate the mean phenotypic value of the character C amongst individuals with Genotype g_i in Environment e_k by x_{ik} , and by f_{ik} their relative frequency in the population. If the effects of the environment and genotype were additive, it would be possible to attribute to each genotype, g_i , an effect α_i and to each environment, e_k , an effect β_k , so that

$$x_{ik} = \mu + \alpha_i + \beta_k. \quad (7)$$

In general, though, this additivity does not correspond to reality; it is necessary to take into account the effects of interaction between genotype and environment. We may write

$$x_{ik} = \mu + \alpha_i + \beta_k + \gamma_{ik}, \quad (8)$$

where γ_{ik} is the interaction effect attributable to Genotype g_i and Environment e_k , and measures the deviation from the additive model for the combination (g_i, e_k) .

In order to determine the α_i , β_k and γ_{ik} using the known values of the x_{ik} , it is necessary to solve a system in which there are more unknowns than equations. A standard procedure consists of looking for a solution which provides small absolute values for the parameters γ_{ik} . The least squares solution minimizes

$$\sum_{i,k} f_{ik} (x_{ik} - \mu - \alpha_i - \beta_k)^2,$$

giving

$$\left. \begin{aligned} \sum_k f_{ik} (x_{ik} - \mu - \alpha_i - \beta_k) &= 0 \\ \sum_i f_{ik} (x_{ik} - \mu - \alpha_i - \beta_k) &= 0 \end{aligned} \right\} \quad (9)$$

These $r + s - 1$ independent equations with $r + s$ unknowns may be solved by imposing the conditions

$$\sum_i f_i \alpha_i = 0, \quad \sum_k f_k \beta_k = 0,$$

where

$$f_{.i} = \sum_k f_{ik}, \quad f_{.k} = \sum_i f_{ik}.$$

Having thus obtained $\{\alpha_i\}$, $\{\beta_k\}$ and $\mu = \sum_{i,k} f_{ik} x_{ik}$, we obtain $\{\gamma_{ik}\}$ by $\gamma_{ik} = x_{ik} - \mu - \alpha_i - \beta_k$.

In this way it is possible to analyse the x_{ik} according to (8), which allows us to write the variance of X :

$$\text{var}(X) = \text{var}(\alpha) + \text{var}(\beta) + \text{var}(\gamma) + 2 \text{cov}(\alpha, \beta) + 2 \text{cov}(\alpha, \gamma) + 2 \text{cov}(\beta, \gamma),$$

where

$\text{var}(\alpha)$ is the part of the total variance due to genetic effects, designated in (6) by $\text{var}_G(X)$;
 $\text{var}(\beta)$ is the part due to effects of the environment, previously designated by $\text{var}_E(X)$;
 $\text{var}(\gamma)$ is the variance of the interaction effects, which we designate by $\text{var}_I(X)$;
 $\text{cov}(\alpha, \beta)$ is the covariance of environmental and genotypic effects;
 $\text{cov}(\alpha, \gamma)$ and $\text{cov}(\beta, \gamma)$ are zero, from (9).

Finally we obtain

$$\text{var}(X) = \text{var}_G(X) + \text{var}_E(X) + \text{var}_I(X) + 2 \text{cov}(G, E). \quad (10)$$

This relation may be written in the form (6) if the last two terms are zero, i.e. if

- (i) there is random association of genotype and environment; then $\text{cov}(G, E) = 0$;
- (ii) all the parameters γ measuring the interaction effects are zero, then $\text{var}_I(X) = 0$.

These conditions are obviously hardly ever fulfilled in reality. Thus the genetic portion of variance described by the ratio (7), and classically designated by the expression 'heritability in the broad sense', or 'coefficient of genetic determination' has no meaning in the general case. The concept clearly has little utility since most of the time it cannot be even defined.

Remark 3. The parameter h_L^2 is sometimes calculated by the mechanical use of a technique which is certainly perfectly correct but, like all techniques, can lead to logical catastrophe when used outside its domain of validity. This technique is the analysis of variance.

As an example of the kind of errors which can be made, suppose that we have a character with three possible genotypes and two categories of environment:

Environment	Frequency	Genotype and frequency		
		g_1 $\frac{1}{3}$	g_2 $\frac{1}{3}$	g_3 $\frac{1}{3}$
e_1	$\frac{1}{2}$	100	140	60
e_2	$\frac{1}{2}$	70	90	110

With the indicated frequencies, $\mu = 100$ and $\text{var}(X) = 750$.

We might try to calculate the 'part' of this variance due to the genotype by the following reasoning. Eliminating the effect due to environment by considering the individuals living in Environment e_1 , we obtain a variance, due only to the genetic effects: $\text{var}_1(X) = 1100$. Doing the same for Environment e_2 gives $\text{var}_2(X) = 200$. The mean of the variance due to genetic effects is thus $\frac{1}{2}(1100 + 200) = 650$, hence $h_L^2 = 0.87$.

But this reasoning could equally be replaced by the following. Eliminating the effect of the

environment by taking the mean of the observed values for the same genotype, we find the values 85, 115 and 85 for g_1 , g_2 and g_3 , respectively. The variance of these values represents the portion of the total variance due to the effect of genotype. This variance is 225, hence $h_L^2 = 0.30$.

These two results differ because the first, 650, obtained as the mean of marginal variances is equal, in our notation, to $\text{var}_G(X) + \text{var}_I(X)$, whereas the second, 225, the variance of the marginal means, is equal to $\text{var}_G(X)$. The term $\text{var}_I(X)$ is, in our example, particularly large; thus, to attempt to obtain a value of 'broad heritability' is particularly absurd.

Remark 4. When the necessary assumptions are valid, h_L^2 represents the contribution of the genetic 'causes' to the variance of the character between individuals (and not to the character itself, an important nuance). We note that this component of variance is defined only for the population studied and can take a completely different value in another population. Thus 'genetic heritability in the broad sense' of a character C does not characterize C as such, but C in the given set of individuals and of environments.

Remark 5. Because of the unrealistic nature of the assumptions necessary for its definition, and because of the difficulty of obtaining accurate estimates, h_L^2 is hardly ever used by geneticists. Its use in other disciplines is largely a sign of a misunderstanding of the true meaning of the parameter. This is the case, notably, in research attempting to measure the 'heritability' of the IQ. If this is a question of 'biometric heritability', i.e. the resemblance between parents and offspring, the results obtained are meaningful. However, the purpose is usually to estimate the 'role' of genes in the variance of the character, and there is total confusion. One of the principal theoreticians in quantitative genetics, Kempthorne (1978), has commented on the inconsistency between the definition of h_L^2 as proposed by geneticists and the conclusions reached by certain psychologists regarding the measure of the heritability of IQ: 'Heritability does not even exist in the human IQ context. Why then argue about the magnitude of an imaginary number?'.

1.3 Genetic Heritability in the Narrow Sense, for the Purpose of Efficiency

In order to develop techniques for improving domestic animal and plant breeds, geneticists try to obtain an accurate knowledge not only of the total genetic variance of a character, but also of the contribution made by each individual gene. Here again it is necessary to make many simplifying assumptions in order to achieve progress. The mode of reasoning resembles that used earlier to separate the effects of the genotype and of the environment. It consists of analysing the effect of Genotype g_1 into components associated with each of the constituent genes, and components corresponding to the interaction between the genes. If, for example, g_1 denotes a group of four genes situated at two loci (1, 2; 3, 4), we can attempt to analyse the effect α_1 of Genotype g_1 on the character by a relation of the form

$$\alpha_1 = \rho_1 + \rho_2 + \rho_3 + \rho_4 + \delta_{12} + \delta_{34} + \epsilon_{12,34},$$

where the ρ_i are the additive effects of the genes, δ_{12} and δ_{34} the interaction effects of genes at the same locus ('dominance'), and $\epsilon_{12,34}$ the interaction effects between genes situated on two different loci ('epistasis').

As in the case of the analysis of genotypic and environmental effects discussed previously, we can seek the decomposition which attributes the largest part of the variance to the terms ρ and the smallest to δ and ϵ [see, for example, Falconer, (1960), or Kempthorne (1957)].

Using many assumptions, we can thus estimate the various ρ , δ and ϵ terms; but, in order to calculate the variance of the α_i we have to take into account the covariances between the terms corresponding to various loci.

A very simple case is that of a character depending on a single locus or on several loci without epistasis (i.e. the ϵ terms are zero) and without linkage (i.e. the genetic structures of the loci are independent). In that case we can write

$$\text{var}_G(X) = 2 \text{var}(\rho) + \text{var}(\delta), \quad (11)$$

$\text{cov}(\rho, \delta)$ being zero by the minimization process δ . The first term of the right-hand side of (11) is the variance, $\text{var}_A(X)$, of the additive effects of the genes on the character, and the second term is the variance, $\text{var}_D(X)$, of the dominance effects.

We then define the 'narrow heritability', denoted by h_N^2 , as

$$h_N^2 = \frac{\text{var}_A(X)}{\text{var}(X)}. \quad (12)$$

The 'narrow heritability' is thus the relative part of the total variance associated with the additive effects of the genes.

Remark 6. To define h_N^2 we are forced to make many assumptions; some doubts may therefore be expressed about the usefulness of this parameter. In fact this usefulness is obvious when the point is to improve certain characteristics in plants or animals. But, in those cases, the situation deliberately created by the breeders is very special. The populations which they wish to improve are usually placed in as homogeneous an environment as possible. Individuals are differentiated mainly by their genetic endowment; in the analysis of their variance according to (10) we can assume that the terms $\text{var}_I(X)$ and $\text{cov}(G, E)$ are zero.

Under these conditions, (12) can be written, in an obvious notation, as

$$h_N^2 = \frac{V_A}{V_A + V_D + V_E}.$$

The 'narrow heritability' characterizes the portion of the variance attributable to additive effects in the observed population. The last phrase is particularly important. Breeders know well that the effect of selection is to modify the genetic structure of the population, and as a consequence, the value of h_N^2 .

Knowledge of this parameter can be very important in guiding the choice of selection techniques used to improve a characteristic, but it cannot throw light on the underlying biological mechanisms at work.

Remark 7. The techniques of plant or animal improvement take into account the influence of kinship relation on the resemblance between individuals. This resemblance is measured by the correlation of the measures X of the character between two individuals having a given parental link; this correlation may be calculated in terms of h_N^2 , by the following reasoning. Consider a set of pairs $(1_i, 2_i)$ all having the same genealogical relationship, and suppose that the character studied satisfies (6) and (7) (i.e. there is no interaction between genotype and environment, and the correlation between the genotype effects α and environment effects β are zero). Then

$$\text{cov}(X_1, X_2) = \text{cov}(\alpha_1, \alpha_2) + \text{cov}(\beta_1, \beta_2).$$

Finally, suppose that the genealogical relationship considered does not entail any environmental similarity between members of a pair. Then $\text{cov}(X_1, X_2)$ reduces to $\text{cov}(\alpha_1, \alpha_2)$. Fisher (1918) showed that this covariance may be calculated from V_A , V_D and the parameters which measure the parental relationship between 1 and 2; in the absence of inbreeding, we can write (using the notation of Jacquard, 1978)

$$\text{cov}(\alpha_1, \alpha_2) = 2\phi_{12}V_A + \Delta_7V_D, \quad (13)$$

where ϕ_{12} is the 'coefficient of kinship' of 1 and 2, and Δ_7 is the probability that each gene of 1 is identical by descent to one, and only one, gene of 2.

The correlation is then

$$r_{1,2} = \frac{\text{cov}(X_1, X_2)}{\text{var}(X)} = \frac{2\phi V_A}{\text{var}(X)} + \frac{\Delta_7 V_D}{\text{var}(X)}. \quad (14)$$

In the case when $\Delta_7 = 0$ (i.e. when one parent of Individual 1 is not related to the parents of 2), (14) reduces to

$$r_{1,2} = 2\phi h_N^2.$$

In this particular case, the correlation between two individuals may thus be expressed as a function of the 'heritability in the narrow sense'.

Remark 8. For a parent-offspring pair $\phi = \frac{1}{2}$ and $\Delta_7 = 0$; hence

$$\text{cov}(P, O) = \frac{1}{2} V_A. \quad (15)$$

But, from a standard result of statistics, the slope of the regression line between offspring and parent is related to their covariance by $k = \text{cov}(P, O)/\text{var}(P)$. From (5) and (15),

$$h^2 = \frac{2\text{cov}(P, O)}{V_P} = \frac{V_A}{V_X} = h_N^2.$$

The parameters measuring the first and the third types of heritability are thus found to be equal. However, this equality, derived by accumulating assumptions, must not be considered to mean that the two concepts represented by h^2 and h_N^2 are equivalent. The 'biometric heritability' h^2 is defined in the space of measures, the space of phenotypes, and characterizes an empirical resemblance. The other quantity, h_N^2 , is defined through a model representing the genetic mechanism of the character being studied. It refers to the space of genotypes, and characterizes the effect of individual factors, the genes, which in reality act only in pairs.

Remark 9. 'Narrow heritability', like the other types of heritability, depends on the population under study. This dependence should be emphasized: the same character may well have a totally different heritability in two populations. To see this, imagine a character completely determined by a locus at which two alleles are segregating, with no environmental effects, and such that the values taken by the three genotypes are

$$\begin{array}{ccc} a_1 a_1 & a_1 a_2 & a_2 a_2 \\ i(= 1.0) & j(= 1.2) & k(= 0.8) \end{array}.$$

Since the environment is assumed to have no influence, $h_L^2 = 1$. The additive and dominance variances are [see, for example, Kempthorne (1957), p. 315]

$$V_A = 2pq\{p(i-j) + q(j-k)\}^2 \text{ and } V_D = p^2 q^2 (i - 2j + k)^2,$$

where p is the frequency of the allele a_1 , and $q = 1 - p$.

Here we obtain

$$V_A = 2p(1-p)(0.4 - 0.6p)^2 \text{ and } V_D = 0.36p^2(1-p)^2.$$

The narrow heritability is thus a function of p . For

$$\begin{array}{ll} p = \frac{1}{3}, & h_N^2 = 0; \\ p = \frac{1}{2}, & h_N^2 = 0.18; \\ p = \frac{2}{3}, & h_N^2 = 0.50. \end{array}$$

Remark 10. In so simple a case, we can obtain values of the two genetic heritabilities of 1 and 0, respectively. This does not arise as an extreme case. Recessive genetic diseases have, by definition, a 'broad heritability' of 1, since they are completely genetically determined, but the 'narrow heritability' is $h_N^2 = 2q/(1 + q)$, where q is the frequency of the gene causing the trait. When $q = 0.02$, as in the case of mucoviscidose in Europe, $h_N^2 = 0.04$. Rare recessive traits have thus a very low 'narrow heritability'. (This is well confirmed by observation since affected offspring almost always have unaffected parents).

These remarks illustrate the danger of the incorrect manipulation of concepts which are quite delicate. The validity of the concept of 'narrow heritability' seems for some to be proved by its efficiency. Animal and plant breeders use the concept continually and their successes can hardly be doubted. However, a less triumphant analysis has been developed in recent years. In August 1976, a symposium of the major theoreticians in this domain met to reflect on 'the deep uncertainty about the quality and utility of available theory of quantitative genetics' (Pollak and Kempthorne, 1977). The most that one can say in 1983 is that the question remains unanswered.

2. Measurement

2.1 Measuring Biometric Heritability

By definition, biometric heritability is the result of an observation: it is the value of the parameter h^2 , which is obtained from the regression function of parents on offspring. If the data consist of (parent, offspring) pairs, h^2 is equal to twice the slope of the regression. If the data consist of (mean-parent, offspring) pairs, h^2 is equal to the slope of the line. It can be shown (Falconer, 1960, p. 181) that the estimated value of h^2 is a random variable with standard deviation approximately $2/T^{\frac{1}{2}}$, where T is the total number of individuals observed.

2.2 The Measurement of Broad Heritability: Comparing Various Types of Twins

Twins can be either monozygotic (MZ), or dizygotic (DZ). Twins of the first kind are the result of a single conception, and represent two expressions of the same genetic inheritance. The differences between such twins can thus be due only to circumstances of life, which we have designated as the 'environment'. In contrast, dizygotic twins come from two distinct conceptions resulting in simultaneous birth; they differ genetically in the same way as ordinary full-sibs (this may be a first approximation; some observations appear to show that dizygotic conceptions are much more frequent than dizygotic births, the latter representing the result of a selection which entails preferential maintenance of those dizygotic twin-pairs which are genetically most similar. We shall neglect this possibility here).

The similarity between members of an MZ pair is naturally larger than within DZ pairs and comparison should allow an estimate of the contribution of the genome to the variability of a character. Classically, we define

$$H = \frac{V_{DZ} - V_{MZ}}{V_{DZ}} \quad (16)$$

(where each V is the variance of differences between twins from the same pair). The parameter H is often designated by the term 'heritability', or 'degree of genetic determination'; in fact the relation between this parameter and the 'broad heritability' previously defined is not clear.

Let (J_{1i}, J_{2i}) be a series of twin-pairs with phenotypic values x_{1i} and x_{2i} for the character C .

We have

$$V_J = \frac{1}{n} \sum_{i=1}^n (x_{1i} - x_{2i})^2 = 2V_x - 2 \operatorname{cov}(x_1, x_2),$$

where V_x is the variance of C calculated by considering one individual from each pair, so

$$H = \frac{\operatorname{cov}_{MZ} - \operatorname{cov}_{DZ}}{V_x - \operatorname{cov}_{DZ}}.$$

The variance V_x can be partitioned, as in (10), into different terms corresponding to effects of the genotype and the environment, and their interaction and covariance. To take account of all these terms makes the calculation intractable; to make progress, we must assume that the interaction and covariance terms are zero, i.e. that (6) and (7) describe the true situation. These hypotheses obviously remove a large part of the significance of the results to be obtained but they are necessary for the development of the argument. We shall further assume that twins, as individuals, are not distinguished from other persons, thus neglecting the specific conditions of their intrauterine life and of their later development.

From (13), the covariances can be calculated with $\phi = \frac{1}{2}$ and $\Delta_7 = 1$ for monozygotes, and $\phi = \frac{1}{4}$ and $\Delta_7 = \frac{1}{4}$ for dizygotes.

Finally, we obtain the classical result [see, for example, Cavalli-Sforza and Bodmer (1971), p. 578)

$$H = \frac{V_A + \frac{1}{2} V_D + \operatorname{cov}_{MZ}(\beta_1, \beta_2) - \operatorname{cov}_{DZ}(\beta_1, \beta_2)}{V_A + \frac{1}{2} V_D + 2V_M - 2 \operatorname{cov}_{DZ}(\beta_1, \beta_2)}. \quad (17)$$

Some supplementary assumptions on the covariances of the environmental effects are necessary to relate H to h_L^2 but they take us still further from reality.

In conclusion it is difficult to regard H as an 'estimator of heritability'; in any case, the heritability in question is the 'broad heritability' which, as we have seen, can only be defined by making unrealistic assumptions and which is thus not very useful. One can wonder why so much trouble is taken to measure a parameter linked to a concept which is usually not definable.

Remark 11. In order to show how much the results depend upon the validity of the assumptions used, suppose that the parents of MZ twins tend to accentuate their resemblance more frequently than the parents of DZ twins. For example, suppose that MZ daughters are dressed in an identical fashion in $\frac{3}{4}$ of families, and DZ twins are dressed identically in $\frac{1}{4}$ of families. Taking C as skirt length, with V as the variance of the character in the population, we would obtain

$$\operatorname{cov}_{MZ} = E(X_1 - \bar{X})(X_2 - \bar{X}) = \frac{3}{4} E(X_1 - \bar{X})^2 = \frac{3}{4} V,$$

$$\operatorname{cov}_{DZ} = \frac{1}{4} V.$$

Since there is no genotypic effect, $V_A = V_D = 0$, and from (17), we obtain

$$H = \frac{\frac{3}{4} V - \frac{1}{4} V}{V - \frac{1}{4} V} = \frac{2}{3}.$$

This simple difference of family attitude results in a higher 'heritability' than exists for certain characters submitted to an evident genetic determinism such as arterial hypertension or cephalic index. Certainly a researcher finding such a large value of H for the length of skirts would not conclude the trait is genetic, since it is known that the genotype could hardly intervene in the processes which determine this character. He would refute the genetic interpretation of H by referring to the mechanism determining the character. We see that,

when the mechanism is unknown, it is not possible to conclude that a high value of H implies a genetic determinism.

Remark 12. With a qualitative character, the twin-pair is not characterized by the difference between two values, but concordance or nonconcordance for the trait. For such a character, we can calculate P_{MZ} and P_{DZ} , the proportions of concordance for each twin-type.

It is customary to calculate the parameter H' defined by Holzinger in 1929:

$$H' = \frac{P_{MZ} - P_{DZ}}{1 - P_{DZ}}.$$

It can be shown (Cavalli-Sforza and Bodmer, 1971, p. 580) that the value of H' can be used, after making a number of assumptions, to calculate upper and lower bounds for h^2_L . The necessary calculations are sufficiently complex to create the illusion that the results have a serious meaning. In fact, this meaning depends not on the complexity of the calculation, but on the validity of the assumptions. In the case of H , we have seen that this validity should be carefully considered.

In conclusion it would be prudent to consider H or H' as specific parameters permitting the comparison of resemblance in MZ and DZ twins, and to avoid any reference to the concept of heritability.

2.3 Measuring Narrow Heritability

Here it is a question of measuring a parameter defined by a model of relationship between the variation of the character under study and the variation of genes and environment. This model is completely represented by (8), but the presence of the interaction terms γ is so troublesome that hardly any progress can be made without the assumption that they are zero: 'there is little we can do about it in practice except ignore it, because . . . there is no practical means of separating the interaction from the other components' (Falconer, 1960, p. 158). In this situation, the simplified model given by (7) must be used.

Under (7), the estimation of h^2_N reduces to the estimation of the additive genetic variance V_A . To achieve this, we use the observed sample covariance of the character in a set of pairs having the same given relationship. The expected value of this covariance is given by (13). From the covariances observed for various parental relationship it is possible to estimate V_A and V_D by making an assumption about the nature of at least two of the covariances between environmental effects.

In animal or plant breeding with artificial selection, or in experiments conducted according to a pre-established protocol, such assumptions may be perfectly founded. For instance, in the case of artificial insemination, the many descendants of one bull are half-siblings, for which it is often possible to take $\text{cov}(\beta_1, \beta_2) = 0$. In some situations the environments of the father and offspring are uncorrelated, but those of the mother and offspring are related. Under these clearly defined conditions, it is thus possible to estimate V_A and hence h^2_N . This permits prediction of the effects of a given selection régime, assuming that the conditions of the estimate are respected.

Here the question is not of measuring some natural characteristic, but only of predicting the consequences of an intervention.

3. Conclusions

The use of a single word, 'heritability', to designate three precise, but totally distinct, concepts, constitutes a trap all the more dangerous since the term is widely used and concerns a process situated at the heart of much biological research, namely the transmission of characters. An

example will show the danger entailed by this imprecision of language: the study of heritability of female fertility.

A general result, related to the concepts discussed here, and which can be derived without any restrictive hypotheses (Jacquard, 1977, p. 98), shows that a stable character can have a nonzero heritability only if it is not correlated to a selective value. We can write

$$\Delta\bar{C} = \frac{h^2}{2\bar{W}} r(V_C V_W)^{\frac{1}{2}},$$

where $\Delta\bar{C}$ is the change of the mean of the character C between two generations due to genetic change, h^2 is the biometric heritability of C , r is the correlation between C and the individual selective value, \bar{W} is the mean selective value, and V_C and V_W are the variances of C and of the selective value, respectively. If $\Delta\bar{C} = 0$, i.e. if we suppose that the character is stable, it is necessary that $h^2 = 0$ or $r = 0$. A character such as female fertility clearly involves $r \neq 0$, hence it must have zero heritability. This conclusion may appear astonishing: it is evident that certain biological characteristics, notably in females, are favourable (or non-favourable) to fertility; these characteristics depend on the genotype and hence are heritable. Everyone can cite cases of lineages in which large families can be traced for generation after generation. But this is a question of individuals, whereas heritability is a collective parameter. That certain females possess genes which are favourable (or nonfavourable), is hardly deniable, but globally, the presence of these genes cannot entail a resemblance between mothers and daughters. It is necessary to repeat that 'heritability' does not characterize the character itself, but the structure of the population in which it is studied.

The need for great rigour exists particularly in the case of research projects which have serious implications for us all; this is the case when psychologists study the 'heritability of intellectual aptitudes'. They should take the precaution of systematically defining in a precise way the sense in which they use the word 'heritability'; they should also state whether the assumptions under which this word can be used hold true in their studies. It is highly probable that most of the time this exercise in rigour would lead them to the conclusion that none of the three parameters proposed by geneticists can be of any use in solving their problems.

ACKNOWLEDGEMENTS

I am greatly indebted to Marck Lathrop and Michel Schiff for the translation from French.

RÉSUMÉ

Le terme 'héritabilité', qui évoque l'image de la transmission des parents aux enfants, est utilisé en biologie pour caractériser la ressemblance d'individus apparentés en terme d'une caractéristique donnée, et pour analyser les causes génétiques et environnementales de cette ressemblance. En fait, il y a trois définitions de l'héritabilité et il est important de les distinguer pour éviter les mauvaises interprétations. Différentes techniques pour mesurer des paramètres associés sont attachées à ces définitions. Une analyse rigoureuse des hypothèses qui permettent l'interprétation des estimations des paramètres est nécessaire pour éviter des conclusions fausses.

REFERENCES

- Cavalli-Sforza, L. L. and Bodmer, W. (1971). *The Genetics of Human Populations*. San Francisco: Freeman.
- Falconer, D. S. (1960). *Introduction to Quantitative Genetics*. London: Oliver and Boyd.
- Fisher, R. A. (1981). On the correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh* 52, 399-433.
- Jacquard, A. (1974). *The Genetic Structure of Populations*. New York: Springer-Verlag.

- Jacquard, A. (1977). *Concepts en Génétique des Populations*. Paris: Masson.
- Kempthorne, O. (1957). *An Introduction to Genetic Statistics*. New York: Wiley.
- Kempthorne, O. (1978). Logical epistemological and statistical aspects of nature-nurture data interpretation. *Biometrics* 34, 1-23.
- Pollack, E. and Kempthorne, O. (1977). In *Proceedings of the International Conference on Quantitative Genetics*. Ames, Iowa: Iowa State University Press.

Received September 1980; revised July 1981 and February 1982

REVUE BIBLIOGRAPHIQUE

LES MÉTHODES DE CALCUL
DU COEFFICIENT D'HÉRITABILITÉ
EN AMÉLIORATION DES PLANTES

F. LE COHEG

*Station d'Amélioration des Plantes,
Centre de Recherches de Rennes
École nationale supérieure agronomique, I. N. R. A.,
35 - Rennes*

RÉSUMÉ

Une revue bibliographique est faite des différentes manières d'estimer le coefficient d'héritabilité d'un caractère en Amélioration des Plantes, en insistant plus particulièrement sur l'intérêt et les limites de chaque méthode.

Mots clés : *héritabilité*.

L'efficacité de la sélection pour des caractères quantitatifs c'est-à-dire la facilité avec laquelle le caractère observé se transmet à la descendance dépend du système de reproduction et des effets génétiques. La sélection sera d'autant plus facile que la part de variabilité due au milieu présente dans la variabilité totale observée sera faible.

Dans ces conditions, l'héritabilité peut être définie de deux manières :

1° Au sens large elle représente la proportion de la variabilité totale observée due à la variabilité génétique. C'est le degré auquel la sélection génotypique peut être basée sur des performances phénotypiques ou la limite de l'amélioration possible par sélection.

2° Au sens restreint elle représente la proportion de la variabilité totale observée qui est due à la partie additive de la variabilité génétique. C'est la fraction héritable de la variabilité observée.

Ces deux aspects de l'héritabilité correspondent à 2 étapes de la sélection :

- choix des reproducteurs (héritabilité au sens large),
- reproduction entre eux des géotypes choisis (héritabilité au sens restreint).

Le facteur le plus important contribuant à l'héritabilité d'un caractère est son degré d'expressivité : les caractères très influencés par le milieu tendent à avoir une faible héritabilité. De même les caractères complexes ont une héritabilité plus faible que les caractères simples.

L'héritabilité est le reflet de la composition génétique de la population étudiée et du milieu dans lequel elle a été mesurée et, en particulier, elle peut changer avec les progrès de la sélection.

Au sens large l'héritabilité s'exprime par le rapport de la variance génotypique à la variance phénotypique :

$$\frac{\sigma_G^2}{\sigma_G^2 + \sigma_E^2}$$

Dans ce sens l'héritabilité donne une mesure globale du déterminisme héréditaire d'un caractère. Comme les interactions milieu \times génotype sont comprises dans σ_G^2 c'est une estimation pour un milieu donné qui a une valeur quand on recherche une adaptation précise. Ce type d'héritabilité comprend la variabilité due à tous les types d'action de gènes : additivité, dominance et épistasie. La variance génotypique est : $\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_I^2$ (1) et l'héritabilité au sens large le rapport :

$$H = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_{GE}^2 + \sigma_e^2}$$

σ_{GE}^2 représentant les interactions du milieu et du génotype et σ_e^2 l'erreur expérimentale.

Au sens restreint, l'héritabilité est le rapport de la variance additive σ_A^2 à la variance phénotypique totale. On utilise alors de préférence le terme h^2 au lieu de H (FALCONER, 1960).

$$h^2 = \frac{\sigma_A^2}{\sigma_G^2 + \sigma_{GE}^2 + \sigma_e^2}$$

La variance additive σ_A^2 est, avec l'effet additif \times additif de la variance d'épistasie, σ_{AA}^2 , la cause principale de ressemblance entre parents et leur descendance. En fécondation croisée et en l'absence d'épistasie, la variance additive mesure le progrès que l'on peut attendre de la sélection pour un caractère donné en utilisant les méthodes de sélection massale. Sa connaissance est d'un grand intérêt car c'est la seule partie de la variance sur laquelle on puisse sélectionner directement. L'héritabilité au sens étroit est d'un intérêt primordial dans la détermination de l'hérédité d'un caractère quantitatif.

I. — HÉRITABILITÉ AU SENS LARGE

Les méthodes de calcul sont différentes suivant que l'estimation de l'héritabilité est faite en prenant comme unité de sélection la plante individuelle ou la parcelle d'un essai.

(1) La variance est représentée par σ^2 ou V suivant la terminologie figurant dans les articles cités.

1. — *Au niveau de la plante individuelle*

Le principe de calcul de l'hérabilité est basé sur le fait que la variance d'une population qui n'est pas en ségrégation (lignée pure, clone, F_1) peut généralement fournir une estimée de la variance due au milieu, l'hypothèse étant que cette variance est la même dans la population qui est en ségrégation et celle qui ne l'est pas.

Ainsi l'hérabilité d'un caractère d'un clone I_0 ayant des descendances autofécondées I_1 est $\frac{V_{I_1} - V_{I_0}}{V_{I_1}}$ (KALTON *et al.*, 1952 ; Mac DONALD *et al.*, 1952).

De même BURTON (1951) à partir de croisements de lignées inbred de Mil a proposé comme estimée de H : $\frac{V_{F_1} - V_{F_2}}{V_{F_1}}$, V_{F_1} étant une estimée de la variance due au milieu.

Chez les plantes autogames la variance due au milieu dans la génération F_2 peut être estimée aussi à partir de la variance F_1 ou encore à partir de :

— la racine carrée du produit des variances V_{P_1} et V_{P_2} , P_1 et P_2 étant les deux parents (MAHMOUD et KRAMER, 1952).

— la racine cubique du produit des variances V_{P_1} , V_{P_2} , V_{F_1} (WEBER et MOORTHY, 1952).

— le tiers de la somme arithmétique V_{P_1} V_{P_2} V_{F_1} (ALLARD, 1960).

Chez les plantes allogames la faible vigueur des lignées inbred les rendent plus sensibles aux effets du milieu et la grande vigueur de la F_1 donne une valeur trop petite pour la variance due au milieu V_E . PURDY et CRANE (1967) ont proposé pour le Mais la formule :

$$V_E = (2V_{F_1} + V_{P_1} + V_{P_2}) 1/4.$$

Dans tous les cas la variance génotypique calculée par ces méthodes est la somme de la variance génétique totale ($V_G + V_D + V_i$) et de la variance due à l'interaction génotype \times milieu. L'hérabilité obtenue est donc une estimée maximale de l'hérabilité véritable.

Dans les conditions précisées par MATHER (1949) (diploïdie, pas d'allèles multiples, de linkage, d'épistasie et d'interaction GE), l'hérabilité au sens large en F_2 est

$$\frac{1/2 D + 1/4 H}{1/2 D + 1/4 H + E}$$

$$\text{en } F_3, \frac{1/2 D + 1/16 H}{1/2 D + 1/16 H + E} \quad \text{en } F_4, \frac{1/2 D + 1/64 H}{1/2 D + 1/64 H + E}$$

D étant la variance due à l'effet additif des gènes.

H étant la variance due à l'effet de dominance des gènes.

A partir de la F_3 la variance de dominance devient négligeable.

2. — *Au niveau de la parcelle expérimentale*

Quand on dispose d'un grand nombre de plantes, les essais agronomiques permettent de réduire considérablement l'interaction du génotype et du milieu et donnent de meilleures estimées de l'hérabilité au sens large.

BURTON et de VANE (1953) ont montré qu'en comparant différents clones de Fétuque dans un essai à N répétitions la quantité « Carrés moyens entre clones » est une estimée de la quantité $V_g + NV_e$, V_g étant la variance génétique totale et V_e le carré moyen de l'erreur.

L'héritabilité au niveau des plantes individuelles est toujours $H_1 = \frac{V_g}{V_g + V_e}$.
Au niveau des N répétitions elle est :

$$H_2 = \frac{V_g}{V_g + \frac{V_e}{N}} \quad \text{avec } H_1 = \frac{H_2}{H_2 + \frac{1 - H_2}{N}}$$

La méthode diminue la variance de l'erreur et donne des estimées de H plus élevées. Elle est indépendante du fait que cette variance soit égale dans les générations en ségrégation et les générations non en ségrégation.

BROOKS (1962) remarque que le terme « diversity » est aussi approprié qu'héritabilité au sens large dans ces comparaisons de clones puisqu'il n'y a pas analyse de générations en ségrégation. Dans ce cas toute la variabilité a son intérêt et pas seulement la variance additive. Il en est de même si l'on compare entre eux des hybrides F_1 destinés à être utilisés directement (HANSON, 1963).

R. E. COMSTOCK et H. F. ROBINSON (1952) ont montré que dans une expérimentation entre lignées à r répétitions faite en l lieux et y années les carrés moyens de l'analyse de la variance sont des estimées de relations définies entre la composante génétique totale due aux différences entre lignées σ_g^2 , les composantes d'interaction σ_{gl}^2 , σ_{gy}^2 , et σ_{gly}^2 et la variance de l'erreur σ_e^2 (méthode des composantes de la variance). A partir des carrés moyens obtenus il est alors facile de déterminer séparément des estimées de chacune des composantes.

JOHNSON, ROBINSON et COMSTOCK (1955) ont montré que la variance phénotypique σ_{ph}^2 entre les moyennes des lignées comparées dans ces conditions est :

$$\sigma_{ph}^2 = \sigma_g^2 + \frac{\sigma_{gl}^2}{l} + \frac{\sigma_{gy}^2}{y} + \frac{\sigma_{gly}^2}{l y} + \frac{\sigma_e^2}{r l y}$$

et l'héritabilité au sens large est $H = \frac{\sigma_g^2}{\sigma_{ph}^2}$

Au niveau des plantes individuelles la variabilité phénotypique est :

$$\sigma_{ph}^2 = \sigma_g^2 + \sigma_{gl}^2 + \sigma_{gy}^2 + \sigma_{gly}^2 + \sigma_e^2$$

(si les lignées comparées sont homozygotes, σ_g^2 comprend seulement la variance additive σ_A^2 et l'effet additif \times additif de la variance d'épistasie σ_{AA}^2).

La formule de BURTON et de VANE est un cas particulier de cette formule générale qui donne une véritable estimée de la variance génotypique, sans les variances d'interaction. Une étude détaillée de ce type d'essai a été faite aussi par HANSON *et al.* (1956) qui l'appliquent à partir de lignées F_1 .

Les travaux de BREESE (1969) ont montré que l'interaction génotype \times milieu ne se manifeste pas d'une manière imprévisible mais que la plus grande partie peut en être expliquée par une régression linéaire entre le comportement individuel de chaque génotype et la moyenne du comportement de l'ensemble des génotypes dans chaque milieu. On peut alors prévoir le comportement relatif des génotypes dans

un grand nombre de milieux. Enfin chaque génotype étant caractérisé par sa moyenne et son comportement au milieu, il est possible, par la méthode des composantes de la variance d'obtenir de meilleures estimées de l'héritabilité.

Différents auteurs ont insisté sur le fait que plusieurs essais sont nécessaires pour avoir une bonne estimée de l'héritabilité au sens large (surtout pour le rendement). Ainsi HANSON (1963) a recommandé deux répétitions dans deux milieux différents pour le Soja, et CHUNG et LIANG (1970) deux lieux et deux années pour le Sorgho.

Un coefficient d'héritabilité au sens large n'est valable que dans le milieu dans lequel il a été déterminé et que pour l'adaptation à ce milieu bien précis. Cette expérimentation poursuivie dans un grand nombre de lieux et au cours d'un grand nombre d'années n'a cependant pas d'autre part beaucoup d'intérêt puisque l'on peut rendre H dans ces conditions aussi proche de 1 que l'on veut.

II. — L'HÉRITABILITÉ AU SENS ÉTROIT

Pour FALCONER et P. ROBINSON (1963) cette définition est la seule qui devrait porter le nom d'héritabilité.

1. — Méthode de la régression

La définition de l'héritabilité au sens étroit se fait par la régression. Dans une population en panmixie la variance additive, en l'absence d'épistasie, est estimée à partir de la covariance parent-descendance (LUSH, 1948) qui est une estimée de la moitié de la variance additive.

L'héritabilité exprimée en termes de covariance (en l'absence d'épistasie) est :

$$h^2 = \frac{2 \text{ cov}_{or}}{\sigma_p^2} = 2b_{or}$$

cov_{or} étant la covariance de la descendance et des parents, σ_p^2 la variance du parent et b_{or} la régression de la descendance sur le parent (en présence d'épistasie $\text{cov}_{or} = 1/2 V_A + 1/4 V_{AA}$ et l'estimation de h^2 est alors biaisée positivement).

Dans les populations allogames, l'héritabilité d'un caractère au sens étroit est donc égale à deux fois la régression de la valeur moyenne de la descendance à la valeur du parent connu. Dans le cas de croisement légitime entre deux parents, l'héritabilité d'un caractère est égale aussi à deux fois la régression de la valeur moyenne de la descendance sur la valeur correspondante de l'un ou l'autre parent, à condition que les variances des deux parents soient du même ordre de grandeur. On écrit : $h^2 = 2b$.

Par contre, la régression de la descendance sur le parent moyen est une estimée directe du rapport $\frac{V_A}{V_p}$ et donc de l'héritabilité au sens étroit (FALCONER, 1960).

Il en est de même pour la régression de la moyenne d'une descendance I_1 obtenue à partir d'un parent I_0 ou, pour les plantes autogames pour la moyenne des plantes

F_2 sur le parent F_1 . Cependant, dans ce cas, la variance de dominance peut introduire un biais dans l'estimation du coefficient d'hérabilité au sens étroit. Dans tous les cas les résultats obtenus ne s'appliquent qu'à la population étudiée dans le milieu envisagé.

MAHMOUD et KRAMER (1951) ont proposé une généralisation de la formule de la régression chez les plantes autogames de la forme :

$$h^2 = \frac{\bar{x}}{\bar{y}} \times b_{yx} \times 100$$

\bar{x} = moyenne de la génération la moins avancée

\bar{y} = moyenne de la génération la plus avancée.

Dans le calcul de la régression, les deux générations comparées étant cultivées dans des milieux différents les effets du milieu tendent à accroître ou à diminuer la variation phénotypique de la descendance par rapport aux parents de sorte que des estimées de H supérieure à 1 peuvent être obtenues.

FREY et HORNER (1957) ont proposé de calculer dans les générations de plantes autogames les régressions à partir de données transformées préalablement en unités d'écart-type. On est donc, dans ces conditions, amené à calculer directement un coefficient de corrélation entre les données initiales (SNEDECOR) au lieu d'un coefficient de régression. Ce coefficient de corrélation nommé b' est toujours inférieur à 1. La méthode de FREY et HORNER ne permet cependant d'éliminer qu'une partie de l'interaction milieu \times génotype présente dans le calcul par la méthode de la régression.

Pour SMITH et KINMAN (1965) la formule $h^2 = b$ (b étant la régression parent-enfant) est bien appropriée pour les plantes autogames si le parent n'est pas inbred, comme dans le cas de la régression F_2/F_1 , mais elle conduit à une surestimation de h^2 si le parent est plus ou moins inbred comme dans le cas de la régression F_3/F_1 .

Avec l'hypothèse que la variance de dominance est négligeable une estimation plus exacte de h^2 est $\frac{b}{2r_{xy}}$, r_{xy} étant le coefficient de parenté (MALÉCOT) entre le parent y et sa descendance x .

r_{xy} est 1/2 pour les générations F_1, F_2 , 3/4 pour les générations F_3, F_4 , 7/8 pour F_5, F_6 , 15/16 pour F_7, F_8 et 31/32 pour F_9, F_{10} .

Cependant, pour OMRAN *et al.* (1968), si le coefficient r_{xy} peut être utilement employé dans les conditions définies par MALÉCOT, son application ne se justifie pas dans les descendance de plantes autogames et $h^2 = b$ est la véritable estimée de l'hérabilité.

Il faut à nouveau remarquer que les chiffres obtenus par la méthode de la régression ne sont valables que pour la population et le milieu considérés. D'autre part cette méthode n'est valable, pour estimer la variance additive, que si les variances de dominance et d'épistasie sont faibles (on obtient ainsi de meilleures estimées dans les générations F_4F_5 que dans les générations F_3F_4). Elle provoque généralement une surestimation de l'hérabilité au sens étroit. Cependant dans le cas où les parents sont consanguins et si la descendance est obtenue en panmixie il peut y avoir surestimation ou sous-estimation (GALLAIS, 1971).

Enfin la méthode de la régression n'est pas exempte du biais introduit par l'interaction GE, même en utilisant la méthode de FREY et HORNER.

Pour COMSTOCK et MOLL (1963), cités par OMRAN, ce biais peut être supprimé en cultivant les deux générations la même année et en répétant l'expérience une seconde année.

Dans ces conditions, pour HANSON (1963), cette méthode et celle des composantes de la variance sont les meilleures pour estimer l'héritabilité.

La méthode de la régression peut s'appliquer, que l'hérédité soit disomique ou tétrasomique. Cependant, dans les cas d'hérédité tétrasomique, même dans le cas d'absence d'épistasie, il peut y avoir biais par existence d'interactions entre 2, 3 et 4 gènes allèles (GALLAIS, 1971).

Pour déterminer l'héritabilité au sens large, l'hypothèse d'absence de compétition n'est pas nécessaire (sauf si elle est estimée à partir de plantes individuelles F_1). Pour l'héritabilité au sens étroit définie par la régression il faut faire cette hypothèse (HANSON, 1963). Mais GRIFFING (1968) a levé cette hypothèse.

ROBINSON *et al.* (1949) et FALCONER (1960) ont donné des formules de calcul de l'écart-type du coefficient de régression.

Dans le cas bien défini des hypothèses de MATHER on a plus précisément :

$$H = b_{F_2/F_1} = \frac{\text{COV}_{F_2/F_1}}{V_{F_2}} = \frac{1/2 D + 1/8 H}{1/2 D + 1/4 H + E}$$

le numérateur étant la covariance génétique entre la F_2 et la moyenne des familles F_1 .

De même :

$$H = b_{F_4/F_2} = \frac{\text{COV}_{F_4/F_2}}{V_{F_4}} = \frac{1/2 D + 1/32 H}{1/2 D + 1/16 H + E}$$

La régression et la corrélation donnent les mêmes résultats si la variance phénotypique des deux générations est égale. En effet :

$$H = r_{F_2/F_1} = \frac{\text{COV}_{F_2/F_1}}{\sqrt{V_{F_1} V_{F_2}}} \text{ (LIANG *et al.*, 1969).}$$

La variance due à la dominance tend à rendre les estimées obtenues par le coefficient de corrélation supérieures à celles obtenues par le coefficient de régression.

2. — Analyse des composantes de la variance génétique

L'analyse des différents types de variance qui composent la variance génétique a été l'objet de nombreux travaux. Les méthodes utilisées à cet effet ont été décrites et discutées par HANSON et ROBINSON (1963) et SPRAGUE (1966).

DUDLEY et MOLL (1969) ont précisé les conditions nécessaires à une bonne estimation de ces variances dans les plantes allogames, les plantes autogames et les plantes à multiplication végétative (elles doivent en particulier être séparées de toute interaction génotype \times milieu). Les variances utilisées dans le calcul de l'héritabilité dépendent du système de sélection et du type de variété utilisés (il y a donc un type d'héritabilité pour chaque système de sélection).

3. — Méthode de WARNER

WARNER (1952) a proposé une méthode de calcul de l'héritabilité au sens étroit, sans calcul de la variance génétique totale, ni de la variance due au milieu. Cette méthode convient à l'étude des plantes autogames et des plantes allogames auto-

fertiles en raison de la difficulté d'estimation de ce dernier type de variance à partir de lignées inbred.

L'héritabilité en F_2 est :

$$\frac{1/2 D}{1/2 D + 1/4 H + E}$$

$1/2 D$ est obtenue par la relation $1/2 D = 2 V_{F_2} - (V_{B_1} + V_{B_2})$, V_{B_1} et V_{B_2} étant les variances phénotypiques des backcross de la F_2 à chaque parent.

Outre les hypothèses de MATHER (absence d'épistasie en particulier) cette méthode requiert l'additivité des effets géniques et l'indépendance entre les moyennes et les variances affectant ces moyennes. Elle requiert en outre que les variances dues au milieu soient du même ordre de grandeur pour la F_2 et les deux backcross.

De grandes populations (plusieurs centaines d'individus dans chaque population) sont nécessaires pour avoir une estimée suffisamment précise de H . Dans ces conditions l'estimée de WARNER est satisfaisante si la variance épistatique est négligeable.

HUET et ÉCOCHARD (1961) ont montré que l'estimée obtenue en F_2 par la méthode de la covariance est égale à la demi-somme de l'estimée au sens large et de l'estimée de WARNER.

4. — Héritabilité réalisée

Un calcul indirect de l'héritabilité peut être fait en considérant que l'héritabilité est le rapport de la réponse à la sélection à la différence en sélection : $h^2 = \frac{R}{S}$ (FALCONER, 1960).

L'héritabilité est alors la part de la différence en sélection que l'on peut effectivement obtenir en identifiant les génotypes supérieurs dans une population de génotypes et mesure un progrès relatif. Sous cet angle l'héritabilité et le gain génétique apparaissent comme deux concepts complémentaires. Cette méthode a été appliquée par NEWELL et EBERHART (1961) et par FREY (1968).

La formule générale de prédiction du progrès génétique est la suivante :

$$G = K \sigma_{ph} H$$

K étant la différence en sélection exprimée en unités d'écart-type, σ_{ph} l'écart-type des variations observées et H l'héritabilité. H doit alors être calculé à partir d'un nombre suffisant d'expérimentations (DUDLEY et MOLL, 1969). Cette conception de l'héritabilité est très utile (HANSON, 1963).

5. — Dans les croisements diallèles, CRUMPACKER et ALLARD (1962) ont défini l'héritabilité au sens étroit comme le rapport :

$$H = \frac{1/4 D}{1/4 D + 1/4 H_1 + 1/4 F + E}$$

D : composante due à l'effet additif des gènes

H_1 : composante due à l'effet de dominance des gènes

F : covariance des deux effets précédents

E : composante due au milieu.

6. — MURTY *et al.* (1962) définissent l'héritabilité au sens étroit comme le rapport $\frac{D}{D+H}$, D et H étant pris en valeur absolue (terminologie de MATHER).

Pour H. F. ROBINSON (1966) il est possible de classer les caractères en trois catégories suivant l'importance de l'héritabilité au sens étroit.

- 1° faible héritabilité (5 à 10 p. 100) : rendement ;
- 2° héritabilité moyenne (10 à 30 p. 100) : composantes du rendement, la taille, les caractères de qualité ;
- 3° héritabilité élevée (30 à 60 p. 100) : les caractères de maturité et de composition chimique.

DISCUSSION

Chez les plantes autogames, l'héritabilité au sens large et au sens étroit, mesurées toutes les deux à chaque génération, se rapprochent progressivement au cours des autofécondations. L'héritabilité est plus élevée dans les premières générations et diminue par la suite. Il est alors plus intéressant de la déterminer dans les générations avancées, car presque toute la variance est additive. Cela n'est cependant pas possible chez la plupart des plantes allogames où la consanguinité est très défavorable (l'estimation de l'héritabilité au sens étroit se fait alors dans les populations par la méthode des composantes de la variance ou la méthode de la régression, mais cette estimation dépend du système de reproduction).

Tous les auteurs insistent pour que soient bien définis le matériel végétal et les conditions de l'expérimentation. Une seule expérimentation ne donne des résultats valables que pour le milieu et la population considérés. Il est nécessaire de bien décrire les caractères utilisés, la population, le milieu et l'unité de sélection : plante individuelle ou parcelle expérimentale (dans ce dernier cas, la dimension des parcelles, la densité du semis, le nombre des répétitions, etc.). D'autre part, il est préférable de déterminer l'héritabilité dans des conditions de productivité agronomique élevées et uniformes qui tendent à donner des estimées plus élevées.

Avec ces conditions, on peut penser que la comparaison de chiffres obtenus par différents auteurs est délicate. Il est cependant possible de comparer différents résultats si, dans chaque expérimentation, l'interaction milieu-génotype a été suffisamment réduite par des essais répétés, par exemple deux années dans deux milieux différents. Avec ces précautions, et bien que les calculs soient faits à partir de formules simplifiées par rapport à la complexité des populations, du milieu et de leur interaction, l'héritabilité constitue un excellent outil pour l'orientation et la prédiction des progrès que l'on peut attendre de la sélection.

Reçu pour publication en mars 1972.

SUMMARY

METHODS OF ESTIMATING THE HERITABILITY COEFFICIENT IN PLANT BREEDING

A review is made of the different methods of estimating the heritability coefficient in plant breeding, laying particular emphasis on the advantages and limits of each method.

RÉFÉRENCES BIBLIOGRAPHIQUES

- ALLARD R. W., 1960. *Principles of plant breeding*. J. Wiley and sons. New York 485 p.
- BRIST E. L., 1969. The measurement and significance of genotype environment interactions in grasses. *Heredity*, **24**, 27-44.
- BROOKS S. N., 1962. Effectiveness of selection within fuggle Hops (*Humulus lupulus*). *Crop Sci.*, **2**, 5-10.
- BURTON G. W., 1951. Quantitative inheritance in pearl millet (*Pennisetum glaucum*). *Agron. J.*, **43**, 409-417.
- BURTON G. W., DE VANE E. H., 1953. Estimating heritabilities in tall Fescue (*Festuca arundinacea*) from replicate clonal material. *Agron. J.*, **45**, 478-481.
- CHUNG J. H., LIANG G. H. L., 1970. Some biometrical studies on nine agronomic traits in Sorghum (*Sorghum bicolor* L.) I. Variance components and Heritability estimates. *Canad. J. Genet. Cytol.*, **12**, 2, 288-296.
- COMSTOCK R. E., ROBINSON H. F., 1952. Genetics parameters, their estimation and significance. *Proc. 6th Intern. Grassland Congr.*, **1**, 284-291.
- COMSTOCK R. E., MOLL R. H., 1963. Genotype-environment interactions. in *Statistical genetics and plant breeding*. Nat. Acad. Sci.-Nat. Res. Council, Publ. 982, 164-196.
- CRUMPACKER D. W., ALLARD R. W., 1962. A diallel cross analysis of heading date in wheat. *Hilgardia*, **32**, 6, 275-318.
- DUDLEY J. W., MOLL R. H., 1963. Interpretation and use of estimates of heritability and genetic variances in plant breeding. *Crop Sci.*, **3**, 257-262.
- FALCONER D. S., 1960. *Introduction to quantitative genetics*. Oliver and Boyd, Edinburg, 363 p.
- FREY K. J., 1968. Expected genetic advances from three simulated selection schemes. *Crop Sci.*, **8**, 235-238.
- FREY K. J., HORNER T., 1957. Heritability in standard units. *Agron. J.*, **49**, 59-62.
- GALLAIS A., 1971. Communication écrite.
- GRIFFING B., 1968. Selection in reference to biological groups. *Aust. J. Biol. Sci.*, **21**, 1163-1170, 1171-1175.
- HANSON W. D., 1963. Heritability. in *Statistical genetics and plant breeding*. Nat. Acad. Sci.-Nat. Res. Council. Publ. 982, 125-140.
- HANSON W. D., ROBINSON H. F., 1963. *Statistical genetics and plant breeding* 623 p. Nat. Acad. Sci.-Nat. Res. Council. Publ. 982. Washington.
- HANSON C. H., ROBINSON H. F., COMSTOCK R. E., 1956. Biometrical studies of yield in segregating populations of Korean Les-pedeza. *Agron. J.*, **48**, 268-272.
- HUET J., ÉCOCHARD R., 1961. Contribution à l'étude de la génétique quantitative chez une plante autogame; le blé. I. Principes fondamentaux. *Ann. Amélior. Plantes*, **11**, 1, 25-59.
- KALTON R. R., SMITH A. G., LEFFEL R. C., 1952. Parent-inbred progeny relationships of selected Orchardgrass clones. *Agron. J.*, **44**, 481-486.
- LIANG G. H. L., WALTER T. L., NICKELL C. D., KOH Y. O., 1969. Heritability estimates and interrelationships among agronomic traits in grain Sorghum. (*Sorghum bicolor* L.) *Canad. J. Genet. Cytol.*, **11**, 199-208.
- LUSH J. L., 1948. *The genetics of populations*. The Iowa State University Press. Ames. Iowa.
- MCDONALD E. D., KALTON R. R., WEISS M. G., 1952. Interrelationships and relative variability among S_1 and open pollination progenies of selected Bromegrass clones. *Agron. J.*, **44**, 20-25.
- MAHMUD I., KRAMER H. H., 1951. Segregation for yield, height and maturity following a soybean cross. *Agron. J.*, **43**, 605-609.
- MALÉCOT G., 1948. *Les Mathématiques de l'hérédité*, MASSON et C^{ie}, Paris. 63 p.
- MATHIE K., 1949. *Biometrical genetics*, Methuen, London. 162 p.
- MURTY B. R., MURTY G. S., PAVATE M. V., 1962. Studies on quantitative inheritance in *Nicotiana tabacum* L. *Der Züchter*, **32**, 8, 361-369.
- NEWELL L. C., EBERHART S. A., 1961. Clone and progeny evaluation in the improvement of Switchgrass, *Panicum virgatum* L. *Crop Sci.*, **1**, 117-121.
- OMMAN A. O., ATKINS I. M., GILMORE E. C., 1968. Heritability of cold hardiness in Flax (*Linum usitatissimum* L.). *Crop Sci.*, **8**, 6, 716-719.
- PURDY J. L., CRANE P. L., 1967. Inheritance of drying rate in mature Corn (*Zea Mays* L.). *Crop Sci.*, **7**, 294-297.
- ROBINSON H. F., COMSTOCK R. E., HARVEY P. H., 1949. Estimates of heritability and the degree of dominance in corn. *Agron. J.*, **41**, 353-359.

- ROBINSON H. F., 1966. Quantitative genetics in relation to breeding on the centennial of Mendelism. *Indian J. Genet. Plant Breed.* **28A**, 171-187.
- ROBINSON P., 1963. Heritability : a second look. in *Statistical genetics and plant breeding*. Nat. Acad. Sci.-Nat. Res. Council-Publ. 982, 609-614.
- SMITH J. D., KINMAN M. L., 1965. The use of parent-offspring regression as an estimator of heritability. *Crop Sci.* **5**, 6, 595-596.
- SNEDCOR G. W., 1956. *Statistical methods*. The Iowa State University Press, Ames, Iowa., 534 p.
- STRAQUE G. F., 1966. Quantitative genetics in Plant improvement. in *Plant breeding*. Frey K. J. (Ed)., 315-354. The Iowa State University Press, Ames, Iowa.
- WARNER J. N., 1952. A method for estimating heritability. *Agron. J.* **44**, 427-430.
- WEBER C. R., MUORTHY B. R., 1952. Heritable and non-heritable relationships and variability of oil content and agronomic characters in the F_2 generation of soybean crosses. *Agron. J.* **44**, 202-209.
-

QUELQUES OUVRAGES DE RÉFÉRENCE

- * DEMARLY Y. 1977. *Génétique et amélioration des plantes*. MASSON.
- * FALCONER D.S. 1974. *Introduction à la génétique quantitative*. Traduction française. MASSON. (3^{ème} édition anglaise datée de 1989 et publiée par LONG-MAN).
- * GALLAIS A. 1990. *Théorie de la sélection en amélioration des plantes*. MASSON.
- * JACQUARD A. 1970. *Structures génétiques des populations*. MASSON.
- * MALECOT G. 1966. *Probabilités et hérédité*. *Travaux et Documents*, Institut National d'Etudes Démographiques. Cahier n47.